

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĐİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĐİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**TÜRKÇE MAGAZİN HABERLERİNDE ADLANDIRILMIŐ VARLIK
TANIMA**

HAZIRLAYAN

MERVE ADAK

YÜKSEK LİSANS TEZİ

ANKARA – 2025

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĐİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĐİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**TÜRKÇE MAGAZİN HABERLERİNDE ADLANDIRILMIŐ VARLIK
TANIMA**

HAZIRLAYAN

MERVE ADAK

YÜKSEK LİSANS TEZİ

TEZ DANIŐMANI

DR. ÖĐR. ÜYESİ MEHMET DİKMEN

ANKARA – 2025

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı çerçevesinde Merve ADAK tarafından hazırlanan bu çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 12 / 08 / 2025

Tez Adı: Türkçe Magazin Haberlerinde Adlandırılmış Varlık Tanıma

Tez Jüri Üyeleri (Unvanı, Adı - Soyadı, Kurumu)

İmza

Dr. Öğr. Üyesi Mehmet DİKMEN, Başkent Üniversitesi

Dr. Öğr. Üyesi İclal ÇETİN TAŞ, Başkent Üniversitesi

Dr. Öğr. Üyesi Begüm MUTLU BİLGE, Ankara Üniversitesi

ONAY

Prof. Dr. Dilek ÇÖKELİLER SERDAROĞLU

Fen Bilimleri Enstitüsü Müdürü

Tarih: ... / 08 / 2025

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

Tarih: 14 / 08 / 2025

Öğrencinin Adı, Soyadı: Merve ADAK

Öğrencinin Numarası: 22210074

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Danışmanın Unvanı/Adı,Soyadı: Dr. Öğr. Üyesi Mehmet DİKMEN

Tez Başlığı: Türkçe Magazin Haberlerinde Adlandırılmış Varlık Tanıma

Yukarıda başlığı belirtilen Yüksek Lisans tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 64 sayfalık kısmına ilişkin, 14 / 08 / 2025 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %1'dir. Uygulanan filtrelemeler:

1. Kaynakça hariç
2. Alıntılar hariç
3. Beş (5) kelimedenden daha az örtüşme içeren metin kısımları hariç

“Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını” inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:

ONAY

Tarih: 14 / 08 / 2025

Öğrenci Danışmanı

Dr. Öğr. Üyesi Mehmet DİKMEN

TEŐEKKÜR

Bu tezin hazırlanmasında bana rehberlik eden ve destek olan tüm deęerli insanlara teőekkür ederim. Özellikle akademik gelişimimde yol gösterici olan hocam Dr. Öğr. Üyesi Mehmet DİKMEN'e içten şükranlarımı sunarım. Eğitim sürecimde katkı sağlayan Başkent Üniversitesi'ndeki tüm hocalarıma da teőekkür ederim.

Her zaman yanımda olan ve sevgileriyle bana güç veren ailem Ayfer ADAK, Sedat ADAK, Alper ADAK ve Tuęba MACAŐ ADAK'a teőekkür ederim. Varlıkları, bu süreci benim için daha anlamlı kıldı.

Hayatımın her anında desteęini esirgemeyen Alperen ÖZDEMİR'e ise bu süreçte verdiği moral ve motivasyon için ayrıca teőekkür ederim.

Son olarak, bu dönemde yanımda olan ve desteęini hissettiren tüm dostlarıma teőekkür ederim.

ÖZET

Merve ADAK

TÜRKÇE MAGAZİN HABERLERİNDE ADLANDIRILMIŞ VARLIK TANIMA

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2025

Bu tez çalışması, Türkçe magazin haberlerinde adlandırılmış varlık tanıma üzerine odaklanmaktadır. Amaç, Türkçenin eklemeli yapısı ve magazin dilinin bağlama dayalı özellikleri karşısında derin öğrenme tabanlı modellerin performansını değerlendirmektir. Bu kapsamda, kişi, tarih, organizasyon, konum ve ilişki gibi on bir varlık türünü içeren, manuel olarak etiketlenmiş alana özgü bir veri kümesi oluşturulmuştur. Etiketleme BIO ve non-BIO formatlarında yapılmış, veriler eğitim ve test kümelerine ayrılmıştır. Üçü Türkçeye özgü, ikisi çok dilli olmak üzere beş farklı dönüştürücü tabanlı model eğitilmiş ve doğruluk, kesinlik, duyarlılık, F1 skoru gibi metriklerin yanı sıra karışıklık matrisi ve t-SNE görselleştirmeleri ile analiz edilmiştir. Karşılaştırmalar, her iki etiketleme sisteminde de en yüksek başarımın mBERT modeli ile elde edildiğini ortaya koymuştur. Çok kelimeli varlıkların tespitinde, kullanılan etiketleme formatı ile modelin yapısal özelliklerinin uyumu belirleyici olmuştur. Ayrıca modellerin karar süreçlerini yorumlayabilmek için LIME tabanlı açıklanabilirlik analizi de gerçekleştirilmiştir. Sonuçlar, derin öğrenme yöntemlerinin Türkçe magazin haberlerinde etkili bir şekilde uygulanabileceğini ve doğru model ile etiketleme formatı uyumu sağlandığında yüksek doğruluk elde edilebileceğini göstermektedir.

ANAHTAR KELİMELEER: Adlandırılmış Varlık Tanıma, Açıklanabilir Yapay Zekâ, Bilgi Çıkarımı, Doğal Dil İşleme

ABSTRACT

Merve ADAK

NAMED ENTITY RECOGNITION IN TURKISH CELEBRITY NEWS

Başkent University Institute of Science

Department of Computer Engineering

2025

This thesis focuses on named entity recognition in Turkish celebrity news. The aim is to evaluate the performance of deep learning-based models against the agglutinative structure of Turkish and the context-dependent nature of magazine language. For this purpose, a domain-specific dataset containing eleven entity types, including person, date, organization, location, and relationship, was manually annotated. Labeling was performed in both BIO and non-BIO formats, and the data was divided into training and test sets. Five transformer-based models, three designed specifically for Turkish and two multilinguals, were trained and analyzed using metrics such as accuracy, precision, recall, and F1 score, as well as confusion matrices and t-SNE visualizations. Comparisons revealed that the highest performance in both labeling systems was achieved by the mBERT model. In the identification of multi-word entities, the compatibility between the chosen labeling format and the structural features of the model proved to be a determining factor. In addition, LIME based explainability analysis was conducted to interpret the decision-making processes of the models. The results demonstrate that deep learning methods can be effectively applied to Turkish celebrity news and that high accuracy can be achieved when the appropriate model and labeling format alignment are ensured.

KEYWORDS: Celebrity News, Explainable AI, Information Extraction, Named Entity Recognition, Natural Language Processing

ÖNSÖZ

Bu tez çalışması, Türkçe magazin haberleri üzerinde adlandırılmış varlık tanıma problemi odağında gerçekleştirilmiştir. Türkçenin yapısal özellikleri ve magazin dilinin bağlama dayalı anlatımı, adlandırılmış varlık tanıma sistemleri açısından çeşitli zorluklar barındırmaktadır. Bu doğrultuda, yalnızca model eğitimi değil, aynı zamanda alana özgü özgün bir veri seti oluşturularak ilerideki çalışmalara kaynaklık etmesi amaçlanmıştır. Tez sürecinde, değerli bilgi ve deneyimleriyle çalışmanın yönünü şekillendirmeye katkı sunan danışmanım Dr. Öğr. Üyesi Mehmet Dikmen'e katkılarından dolayı kendisine teşekkür ederim. Derin öğrenme tabanlı çeşitli modellerle gerçekleştirilen deneyler hem görsel hem de istatistiksel analizlerle desteklenmiştir. Ayrıca model kararlarının açıklanabilirliğine özel önem verilmiştir. Bu çalışmanın, Türkçe doğal dil işleme ve adlandırılmış varlık tanıma alanına katkı sağlaması ve sonraki araştırmalara zemin oluşturması en büyük temennimdir.

İÇİNDEKİLER

TEŞEKKÜR.....	i
ÖZET	ii
ABSTRACT	iii
ÖNSÖZ.....	iv
İÇİNDEKİLER.....	v
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ	ix
SİMGELER VE KISALTMALAR LİSTESİ	xi
1. GİRİŞ.....	1
1.1. Problem Durumu.....	1
1.2. Araştırma Sorusu ve Hipotez	2
1.3. Çalışmanın Amacı ve Kapsamı	2
1.4. Araştırmanın Önemi	3
2. LİTERATÜR İNCELEMESİ.....	4
2.1. Adlandırılmış Varlık Tanıma.....	4
2.2. Türkçe NER Üzerine Yapılmış Çalışmalar.....	4
2.3. Uygulama Alanları ve Gelecek Perspektifi	7
3. TEMEL KAVRAMLAR VE YÖNTEMLER	8
3.1. Doğal Dil İşleme (Natural Language Processing – NLP).....	8
3.2. Dönüştürücü Tabanlı Derin Öğrenme Modelleri.....	9
3.2.1. BERTurk.....	10
3.2.2. mBERT	13
3.2.3. XLM-R	16
3.2.4. DistilBERTurk.....	18
3.2.5. ConvBERTurk.....	21
3.3. Açıklanabilir Yapay Zekâ (Explainable AI – XAI).....	23

3.4. Değerlendirme Metrikleri.....	23
3.4.1. Kesinlik, Duyarlılık ve F1 Skoru.....	24
3.4.2. Makro ve Ağırlıklı Değerlendirme Metrikleri.....	24
3.4.3. Karışıklık Matrisi	24
3.4.4. ROC-AUC ve PR Eğrileri.....	24
3.4.5. t-SNE ile Görselleştirme	24
3.4.6. Varlık Seviyesi ve Sözcük Birimi Seviyesi Değerlendirme	25
3.5. Veri Seti	25
3.6. Deneysel Ortam ve Sistem Gereksinimleri	26
4. DENEY ÇALIŞMALARI	27
4.1. Deneysel Kurulum ile Eğitim ve Değerlendirme Süreci	27
4.2. Model Bazlı Performans Sonuçları.....	28
4.2.1. BERTurk (non-BIO)	28
4.2.2. mBERT (non-BIO)	31
4.2.3. XLM-R (non-BIO).....	33
4.2.4. DistilBERTurk (non-BIO).....	36
4.2.5. ConvBERTurk (non-BIO).....	38
4.2.6. BERTurk (BIO).....	40
4.2.7. mBERT (BIO).....	43
4.2.8. XLM-R (BIO)	45
4.2.9. DistilBERTurk (BIO).....	47
4.2.10. ConvBERTurk (BIO).....	49
4.3. Modeller Arası Karşılaştırmalı Performans Analizi.....	51
4.4. LIME Yöntemi ile Model Açıklanabilirliğinin Değerlendirilmesi.....	54
4.4.1. non-BIO Etiketleme Üzerine LIME Analizi	55
4.4.2. BIO Etiketleme Üzerine LIME Analizi	58
5. SONUÇ VE ÖNERİLER	62

5.1. Sonuç.....	62
5.2. Öneriler	63
KAYNAKLAR.....	65

EKLER

- EK 1: Varlık İsim Tanıma Tezleri Örnekleri**
- EK 2: Modellerin Karşılaştırmalı Özellikleri**
- EK 3: Etiketleme Formatlarına Göre Sınıf Dağılımı**
- EK 4: BERTurk Eğitim Sonuçları (non-BIO)**
- EK 5: mBERT Eğitim Sonuçları (non-BIO)**
- EK 6: XLM-R Eğitim Sonuçları (non-BIO)**
- EK 7: DistilBERTurk Eğitim Sonuçları (non-BIO)**
- EK 8: ConvBERTurk Eğitim Sonuçları (non-BIO)**
- EK 9: BERTurk Eğitim Sonuçları (BIO)**
- EK 10: mBERT Eğitim Sonuçları (BIO)**
- EK 11: XLM-R Eğitim Sonuçları (BIO)**
- EK 12: DistilBERTurk Eğitim Sonuçları (BIO)**
- EK 13: ConvBERTurk Eğitim Sonuçları (BIO)**
- EK 14: non-BIO Etiketleme Formatında Modellerin Sınıf Bazlı F1 Skorları**
- EK 15: BIO Etiketleme Formatında Modellerin Sınıf Bazlı F1 Skorları**

TABLULAR LİSTESİ

	Sayfa
Tablo 4.1. Modellerin Performans Ölçütleri.....	52
Tablo 4.2. Modellerin LIME Skor Ortalamaları (non-BIO).....	56
Tablo 4.3. Modellerin LIME Skor Ortalamaları (BIO).....	59

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. BERTurk Model Mimarisi.....	12
Şekil 3.2. mBERT Model Mimarisi.....	15
Şekil 3.3. XLM-R Model Mimarisi.....	17
Şekil 3.4. DistilBERTurk Model Mimarisi.....	20
Şekil 3.5. ConvBERTurk Model Mimarisi.....	22
Şekil 4.1. Normalize Edilmiş Karışıklık Matrisi (BERTurk, non-BIO).....	29
Şekil 4.2. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (BERTurk, non-BIO).....	30
Şekil 4.3. Normalize Edilmiş Karışıklık Matrisi (mBERT, non-BIO).....	32
Şekil 4.4. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (mBERT, non-BIO).....	33
Şekil 4.5. Normalize Edilmiş Karışıklık Matrisi (XLM-R, non-BIO).....	34
Şekil 4.6. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (XLM-R, non-BIO).....	35
Şekil 4.7. Normalize Edilmiş Karışıklık Matrisi (DistilBERTurk, non-BIO).....	37
Şekil 4.8. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (DistilBERTurk, non-BIO).....	38
Şekil 4.9. Normalize Edilmiş Karışıklık Matrisi (ConvBERTurk, non-BIO).....	39
Şekil 4.10. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (ConvBERTurk, non-BIO).....	40
Şekil 4.11. Normalize Edilmiş Karışıklık Matrisi (BERTurk, BIO).....	42
Şekil 4.12. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (BERTurk, BIO).....	43
Şekil 4.13. Normalize Edilmiş Karışıklık Matrisi (mBERT, BIO).....	44
Şekil 4.14. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (mBERT, BIO).....	45

Şekil 4.15. Normalize Edilmiş Karışıklık Matrisi (XLM-R, BIO).....	46
Şekil 4.16. Etiket Bazlı t-SNE Kümelenme Görselleştirme (XLM-R, BIO).....	47
Şekil 4.17. Normalize Edilmiş Karışıklık Matrisi (DistilBERTTurk, BIO).....	48
Şekil 4.18. Etiket Bazlı t-SNE Kümelenme Görselleştirme (DistilBERTTurk, BIO).....	49
Şekil 4.19. Normalize Edilmiş Karışıklık Matrisi (ConvBERTTurk, BIO).....	50
Şekil 4.20. Etiket Bazlı t-SNE Kümelenme Görselleştirme (ConvBERTTurk, BIO).....	51
Şekil 4.21. Etiket Türüne Göre LIME Skorları (mBERT, non-BIO).....	57
Şekil 4.22. PERSON Etiketini İçin Örnek LIME Açıklaması (mBERT, non-BIO).....	58
Şekil 4.23. Etiket Türüne Göre LIME Skorları (mBERT, BIO).....	60
Şekil 4.24. B-PERSON Etiketini İçin Örnek LIME Açıklaması (mBERT, BIO).....	61

SİMGELER VE KISALTMALAR LİSTESİ

AI	Artificial Intelligence (Yapay Zekâ)
BERT	Bidirectional Encoder Representations from Transformers
BiLSTM	Bidirectional Long Short-Term Memory (Çift Yönlü Uzun Kısa Süreli Hafıza)
BIO	Begin-Inside-Outside (Başlangıç-İç-Bitiş)
CoNLL	Conference on Natural Language Learning (Doğal Dil Öğrenimi Konferansı)
CRF	Conditional Random Field (Koşullu Rastgele Alan)
GELU	Gaussian Error Linear Unit (Gauss Hatalı Doğrusal Birimi)
GRU	Gated Recurrent Unit (Kapılı Tekrarlayan Birim)
HMM	Hidden Markov Model (Gizli Markov Modeli)
IOBES	Inside-Outside-Beginning-End-Single (Bir Varlık İsmi Etiketleme Formatı)
LIME	Local Interpretable Model-Agnostic Explanations (Yerel Yorumlanabilir Modelden Bağımsız Açıklamalar)
LSTM	Long Short-Term Memory (Uzun Kısa Süreli Hafıza)
ML	Machine Learning (Makine Öğrenmesi)
MLM	Masked Language Modeling (Maskeleye Temelli Dil Modelleme)
MUC	Message Understanding Conference (Mesaj Anlama Konferansı)
NER	Named Entity Recognition (Varlık İsmi Tanıma)
NSP	Next Sentence Prediction (Sonraki Cümleyi Tahmin Etme)
NLP	Natural Language Processing (Doğal Dil İşleme)
SHAP	SHapley Additive exPlanations (Shapley Katkılı Açıklamalar)
SVM	Support Vector Machine (Destek Vektör Makinesi)
VİT	Varlık İsmi Tanıma
XAI	Explainable Artificial Intelligence (Açıklanabilir Yapay Zekâ)

1. GİRİŞ

Doğal dil işleme (Natural Language Processing – NLP) alanında son yıllarda yaşanan gelişmeler, metinlerde yer alan anlamlı öğelerin otomatik olarak tanımlanmasını mümkün kılmıştır [1]. Bu doğrultuda, Adlandırılmış Varlık Tanıma (NER) sistemleri; kişi, yer, organizasyon ve tarih gibi varlık türlerini metin içerisinde tespit ederek bilgi çıkarımı, özetleme, makine çevirisi ve soru-cevap sistemleri gibi pek çok NLP uygulamasında kritik bir rol oynamaktadır [1], [2].

Ancak Türkçenin eklemeli yapısı, morfolojik zenginliği ve sözdizimsel esnekliği, özellikle biçimsel olmayan metinlerde NER sistemlerinin performansını olumsuz yönde etkileyebilmektedir [3]. Çelikkaya, Torunoğlu ve Eryiğit'in [4] çalışmasında; sosyal medya ve konuşma diline dayalı veri kümelerinde, yazım kurallarının ihlali, büyük harf eksikliği ve yabancı isimlere Türkçe ek getirilmesi gibi sorunların, adlandırılmış varlıkların doğru tanımlanmasını güçleştirdiği vurgulanmıştır.

Bu tez çalışmasında, Türkçe magazin haberleri gibi gündelik, argo ve bağlamsal öğeler içeren metinlerde NER sistemlerinin başarımı, derin öğrenme tabanlı farklı modelleme yaklaşımları aracılığıyla analiz edilmiştir. Ayrıca, bu tür içeriklerin varlık tanıma sistemleri üzerindeki etkisi sistematik olarak değerlendirilmiştir.

1.1. Problem Durumu

NER, yapılandırılmamış metinlerden bilgi çıkarımı sürecinin merkezinde yer alan temel görevlerden biridir [1]. Ancak Türkçenin dilsel özellikleri, bu sistemler açısından önemli zorluklar barındırmaktadır. Özellikle kelime biçimlerinin çeşitliliği, eklemeli yapı nedeniyle artmaktadır. Bu durum veride seyrekliğe neden olarak modellerin öğrenme süreçlerini güçleştirmektedir [3].

Bu zorluklar, sosyal medya, forum ve benzeri kullanıcı kaynaklı metinlerde daha belirgin hale gelmektedir. Bu tür metinlerde, yazım kurallarına uyulmaması, özel adların kesme işareti olmadan yazılması, harf tekrarları, karakter bozulmaları (ör. “Justin Bieber”, “Aydiiiiin”) ve yabancı sözcüklere Türkçe ek getirilmesi gibi biçimsel olmayan yapılar sıkça görülmektedir [4].

Küçük ve arkadaşlarının çalışmasında, haber metinleri ile karşılaştırıldığında, Twitter verisi üzerinde NER sistemlerinin F1 skorlarının 0,91'den 0,19'a kadar gerilediği rapor

edilmiştir [5]. Bu durum, biçimsel olmayan içeriklerin NER sistemleri için ne denli zorlayıcı olduğunu açıkça göstermektedir.

Magazin haberleri ise, sıkça takma adlar, metaforik ifadeler, argo kullanımı ve bağlama bağımlı anlatımlar içermesi nedeniyle, varlık türlerinin doğru sınıflandırılabilmesi açısından özel olarak uyarlanmış modellere ve alana özgü etiketlenmiş veri kümelerine ihtiyaç duymaktadır [4], [6].

1.2. Araştırma Sorusu ve Hipotez

Türkçe tweet'ler üzerinde yapılan çalışmalar; argo ifadeler, yazım bozuklukları, çok anlamlılık ve gündelik dile özgü yapılar gibi unsurların, NER sistemlerinin başarımını ciddi ölçüde zorlaştırdığını göstermiştir [5].

Bu tezde, Türkçe magazin haberlerinde yer alan adlandırılmış varlıkların hangi düzeyde doğru tanımlanabildiği, hangi modelleme stratejilerinin daha yüksek performans sunduğu ve bu içeriklerin NER sistemlerinin genel başarımına olan etkileri incelenmiştir. Bu kapsamda aşağıdaki temel araştırma sorusu belirlenmiştir:

Araştırma Sorusu:

Türkçe magazin haberlerinde geçen varlık isimleri, önceden eğitilmiş derin öğrenme tabanlı modeller kullanılarak hangi düzeyde doğrulukla tespit edilip sınıflandırılabilir?

Hipotez:

Dönüştürücü tabanlı dil modelleri, Türkçe magazin haberlerinde geçen varlık isimlerini yüksek doğrulukla tanıyıp doğru sınıflara ayırabilir.

1.3. Çalışmanın Amacı ve Kapsamı

Bu çalışmanın temel amacı, Türkçe magazin haberlerinden oluşturulmuş, elle etiketlenmiş bir veri kümesi üzerinde, farklı dönüştürücü tabanlı modellerin NER performanslarını karşılaştırmalı olarak değerlendirmektir. Bu doğrultuda hem Türkçeye özgü olarak eğitilmiş hem de çok dilli önceden eğitilmiş toplam beş farklı model kullanılmıştır. Her biri BIO ve non-BIO etiketleme şemaları altında olmak üzere on farklı deneysel kurgu gerçekleştirilmiştir.

Sadece model çıktılarının nicel ölçütlerle değerlendirilmesiyle sınırlı kalınmamış, aynı zamanda açıklanabilirlik, görsel analiz incelemeleri de kapsam dâhilinde ele alınmıştır.

1.4. Arařtırmanın Önemi

Türkçede NER üzerine gerçekleştirilen çalışmalar genellikle haber metinleri, resmî belgeler ve sosyal medya içerikleriyle sınırlı kalmaktadır. Fakat magazin haberleri, kendine özgü bağlamdan kopuk anlatımı, esnek dil yapısı ve mecazlı ifadeleri nedeniyle farklı düzeyde zorluklar içermektedir.

Bu çalışma, literatürde sınırlı yer bulan bu tür içerikler üzerine odaklanarak hem veri seti hem de modelleme düzeyinde özgün katkılar sunmayı amaçlamaktadır. Bu doğrultuda oluşturulan alana özgü veri kümesi üzerinde, tek dilli ve çok dilli dil modelleri ile gerçekleştirilen deneyler, NER sistemlerinin başarımını farklı açılardan değerlendirme fırsatı sunmuştur.

İkinci bölümde, NER sistemlerine ve Türkçe üzerine yapılan çalışmalara yönelik literatür sunulmuştur. Üçüncü bölümde, yöntemler, değerlendirme metrikleri ve veri setine ilişkin detaylara yer verilmiştir. Dördüncü bölümde, deneysel kurulum ve elde edilen sonuçlar sunulmuş; beşinci ve son bölümde ise genel çıkarımlar ve öneriler paylaşılmıştır.

2. LİTERATÜR İNCELEMESİ

Doğal Dil İşleme (NLP), metinlerin anlamsal ve yapısal analizini mümkün kılarak bilgi çıkarımı, metin özetleme, makine çevirisi ve soru-cevap sistemleri gibi çok sayıda uygulamaya temel teşkil etmektedir. Bu uygulamalar içerisinde NER, metin içerisindeki kişi, yer, organizasyon, zaman ve benzeri varlıkların tanımlanarak etiketlenmesini hedefleyen temel görevlerden biridir [1], [2]. NER, yapılandırılmamış metinleri anlamlı bilgi birimlerine ayırarak metinsel verilerin yapılandırılmasına katkı sağlar ve birçok NLP uygulamasının ilk adımını oluşturmaktadır [3].

2.1. Adlandırılmış Varlık Tanıma

NER sistemleri, başlangıçta yalnızca kişi, yer ve organizasyon gibi temel varlık türlerinin tespitine odaklanmıştır. Zaman içerisinde bu kapsam, para birimleri, yüzde ifadeleri, sanat eserleri, sosyal medya unsurları ve sağlık verileri gibi daha farklı kategorilerle genişletilmiştir [1], [4]. Erken dönem çalışmalarda, dil bilgisel kurallara ve kalıp eşleştirmeye dayalı kural tabanlı yöntemler yaygın olarak kullanılmış, ancak bu yöntemler yüksek doğruluk sağlamak için yoğun insan emeği gerektirmiştir.

Daha sonraki dönemde, öğrenme temelli istatistiksel yaklaşımlar, örneğin Gizli Markov Modeli (HMM) ve Koşullu Rastgele Alanlar (CRF), kural tabanlı sistemlerin yerini almaya başlamıştır [5]. Son yıllarda ise derin öğrenme alanındaki gelişmelerle birlikte, çift yönlü uzun-kısa süreli bellek ağları (BiLSTM) ve CRF gibi geleneksel yöntemlerin yanı sıra, bağlamsal bilgi sağlayan dönüştürücü tabanlı mimariler (BERT, RoBERTa, XLM-R gibi) öne çıkmıştır [6], [7].

Büyük ölçekli veri kümeleri üzerinde önceden eğitilmiş olan bu modeller, dilin bağlamını daha etkin şekilde öğrenebilmekte; çok anlamlı sözcüklerin doğru yorumlanması ve karmaşık dilbilgisel yapıların çözümünde önemli başarılar sergilemektedir [8].

2.2. Türkçe NER Üzerine Yapılmış Çalışmalar

Türkçede NER çalışmaları, diğer birçok dile kıyasla daha geç başlamış olmakla birlikte, zamanla zenginleşmiş ve yöntemsel çeşitliliğe kavuşmuştur. Erken dönem çalışmalarda dil bağımsız sistemler üzerinden yapılan testlerle sınırlı kalan bu alan, günümüzde Türkçeye özel veri kümeleri, model mimarileri ve morfolojik analizlere dayalı çözümlerle gelişmektedir.

İlk örneklerden biri olan Cucerzan ve Yarowsky'nin çalışması, Türkçe dâhil olmak üzere beş farklı dil üzerinde test edilmiş ve Türkçe için 0,5304 F1 skoru elde edilmiştir [8]. Bu sonuç, Türkçenin yapısal özelliklerinin NER sistemleri için ciddi bir zorluk teşkil ettiğini ortaya koymuştur. Türkçenin sondan eklemeli yapısı, zengin morfolojik ek ve kök üretimi ile esnek sözdizimi gibi özellikleri genel dil modellerinin doğrudan Türkçeye uygulanabilirliğini sınırlamaktadır. Bu nedenle Türkçeye özgü yöntem ve özellik çıkarımlarına olan ihtiyacı artırmaktadır.

Bayraktar, finansal metinler üzerinde gerçekleştirdiği çalışmada, yerel dil bilgisi kurallarını temel alarak kişi adlarını tanıma görevini başarıyla yerine getirmiş ve bu tür sektör odaklı uygulamalarda kural tabanlı sistemlerin hâlâ geçerli olabileceğini göstermiştir [9]. Yavuz ise Bayes tabanlı öğrenme yöntemlerini Türkçeye uygulayarak sözcük uzunluğu, büyük/küçük harf kullanımı gibi temel parametrelerle hibrit bir model oluşturmuştur. Bu yaklaşım doğruluk oranlarının yükseltilmesini sağlamıştır [10].

Özkaya'nın e-posta metinleri üzerinde gerçekleştirdiği çalışmada CRF algoritması kullanılarak kişi, yer ve organizasyon varlıklarının etiketlenmesi hedeflenmiştir. Kural tabanlı sistemlerle CRF'nin birleştirilmesiyle daha dengeli bir başarı elde edilmiştir [11]. Sosyal medya platformlarında dilin yapısal bozuklukları ile başa çıkmak üzere Eken, kelime başı ve sonu harflerine dayalı bir model önererek 0,06'lık başarı artışı sağlamıştır [12].

Öztürkmenoğlu, CRF ve HMM gibi geleneksel istatistiksel modelleri karşılaştırarak özellikle BIO formatlı etiketleme ve belirli pencere boyutlarıyla en iyi performansın CRF modeliyle elde edildiğini göstermiştir [13]. Akkaya, sosyal medya metinleri için LSTM ve transfer öğrenme tekniklerini birleştirmiş ve 0,6572 F1 puanı ile dilin morfolojik karmaşıklıklarına rağmen tatmin edici sonuçlar elde etmiştir [14].

Güneş'in çalışması, BiLSTM ve DBLSTM mimarileri üzerinden yürütülmüş ve ENAMEX dışında kalan varlık türleri (örneğin TIMEX, NUMEX) de dahil edilerek hibrit bir yapı önerilmiştir. Bu modelle 0,9369 F1 puanı elde edilmiştir [15]. Keskin ise Wikipedia üzerinden otomatik veri etiketleme sistemi kurmuş, CRF ile TimeDistributed katmanlı softmax modellerini karşılaştırmış ve 0,9486 gibi yüksek bir başarı elde etmiştir [16].

Eşref, kelime içi yapının karakter düzeyinde analizini mümkün kılan LSTM-CRF mimarisıyla dikkat mekanizmalarının entegre edildiği bir sistem geliştirmiştir. Bu sistem, dilin morfolojik yapısını doğrudan yakalayabildiği için daha yüksek başarı sağlamıştır [17]. Çoban'ın sosyal medya verileri üzerinden gerçekleştirdiği analiz ise NER bilgisinin gizlilik ihlallerine etkisini incelemiş ve sosyal medya odaklı NLP çalışmalarına toplumsal bir boyut kazandırmıştır [18].

Dinç, finansal metinler için iki yeni veri kümesi geliştirmiş ve bu veri kümeleri üzerinde derin öğrenme temelli modelleri test ederek Türkçe NER başarımını artırmıştır [19]. Gencer, İngilizce ve Türkçeyi karşılaştırmalı olarak ele almış ve LSTM ve GRU (Kapalı Tekrarlayan Birim) gibi zaman temelli modellerin Türkçe gibi morfolojik açıdan karmaşık dillerde daha stabil sonuçlar verdiğini göstermiştir [20].

Çarık, Wikipedia ve diğer dış bilgi kaynaklarını NER sürecine entegre eden ELBERT ve ELSemantic modellerini geliştirmiş ve bağlamı zayıf kısa metinlerde bu modellerin başarısını ortaya koymuştur [21]. Gürcan, kelime gömme teknikleri üzerine yaptığı karşılaştırmalı çalışmada, BERT'in semantik anlam derinliği açısından üstün olduğunu, FastText'in ise morfolojik genelleme kapasitesi ile öne çıktığını belirtmiştir [22]. Altıntaş, Türkçeye özel bir ayrıştırıcı geliştirerek BERT-TR modelini eğitmiş ve yerli bir ayrıştırıcının doğruluk üzerindeki etkisini deneysel olarak ortaya koymuştur [23].

Son yıllarda yapılan çalışmalar arasında, Khudoyberdieva'nın Türkçe dili için geliştirdiği 23 alt kategoriden oluşan detaylı etiketleme şeması dikkat çekmektedir. Bu çalışmada, dönüştürücü tabanlı derin öğrenme modelleriyle gerçekleştirilen deneylerde; genel varlık türleri için 93,34, alt kategoriler için ise 79,82 F1 skoru elde edilmiştir [28].

İncidelen'in hukuk metinleri üzerine gerçekleştirdiği çalışmada ise alan adaptasyonu konusuna odaklanılmıştır. BERTurk ve mBERT modellerinin hukuk terminolojisine uyarlanarak yaklaşık 0,01 ila 0,03 arasında F1 artışı sağlandığı gözlemlenmiştir [29].

Türkçe spor haberleri üzerine yapılan Kılıç'ın tezi, bu çalışmaya en çok referans olan işlerden biridir. Bu tezde, BIO ve non-BIO etiketleme biçimleri ayrı ayrı değerlendirilmiştir. LIME açıklamaları ile model kararları şeffaflaştırılmış ve NER sistemlerinin yalnızca başarımı değil, açıklanabilirliği de derinlemesine incelenmiştir [24].

Bu çalışmada da Kılıç'ın tezinden ilhamla, Türkçe magazin haberlerine özgü bir veri seti oluşturulmuş, derin öğrenme tabanlı modellerin performansı karşılaştırmalı olarak test edilmiş ve elde edilen sonuçlar LIME açıklamaları ile yorumlanmıştır. Ek 1'de sunulan seçilmiş tez örnekleri, Türkçe NER literatürünün yıllar içindeki çeşitlenmesine dair sınırlı fakat anlamlı bir görünüm sunmaktadır.

Sonuç olarak, Türkçe NER araştırmalarında kural tabanlı sistemlerden derin öğrenme temelli modellere geçiş süreci açıkça gözlemlenmektedir. Bu süreçte açıklanabilirlik, veri kümesi kalitesi, morfolojik analiz yetenekleri ve sektör odaklı çözümler (örneğin finans, sosyal medya, hukuk) araştırmaların yönünü belirlemektedir. Geliştirilen sistemlerin sadece doğruluk oranları değil, aynı zamanda hangi bağlamda ve neden doğru kararlar verdiğinin anlaşılabilir olması da güncel çalışmalarda kritik bir gereklilik haline gelmiştir.

2.3. Uygulama Alanları ve Gelecek Perspektifi

NER, doğal dil işleme sistemlerinde yapılandırılmamış metinlerin işlenerek anlamlı, yapılandırılmış bilgiye dönüştürülmesinde temel bir rol üstlenmektedir. Kişi, yer, organizasyon, tarih gibi varlıkları metin içerisinde ayıklayarak bilgi çıkarımı, içerik filtreleme, otomatik etiketleme ve belge sınıflandırma gibi çok sayıda NLP uygulamasının ilk adımını oluşturmaktadır [1], [25], [26].

Sağlık, hukuk, finans, haber, eğitim, sosyal medya ve kamu hizmetleri gibi çok çeşitli sektörlerde NER sistemleri; bilgiye hızlı erişim, otomatik karar destek sistemleri, bilgi tabanı oluşturma, ilişki çıkarımı ve risk analizi gibi işlemlerde büyük katkılar sağlamaktadır. Özellikle tıbbi belgelerde hastalık isimlerinin, ilaçların ve prosedürlerin otomatik çıkarımı; finansal haberlerde kurum, kişi ve ekonomik terimlerin belirlenmesi; sosyal medya içeriklerinde gizlilik ihlallerinin izlenmesi gibi uygulama örnekleri dikkat çekmektedir [27], [28].

Bu bağlamda, sektörel bazlı veri kümelerinin oluşturulması ve bu alanlara özgü modellerin eğitilmesi giderek önem kazanmaktadır. Khudoyberdieva [24] tarafından yürütülen çalışmada, Türkçe varlıkları alt kategorilerle ayrıntılandıran özgün bir etiketleme yaklaşımı önerilmiştir. Detaylandırılmış sınıflandırmanın bilgi çıkarımı sürecini daha anlamlı kıldığı vurgulanmıştır. İncidelen [25] ise düşük kaynaklı diller için alan adaptasyonu kavramını merkeze alarak hukuk metinlerinde NER başarımını değerlendirmiş ve alan uyarlaması yapılan modellerin standart modellere göre daha başarılı çıktılar ürettiğini ortaya koymuştur.

Gelecek perspektifi açısından bakıldığında, NER sistemlerinin sadece doğruluk oranlarıyla değil, aynı zamanda açıklanabilirlik ve güvenilirlik ölçütleriyle de değerlendirilmesi gerekmektedir. Açıklanabilir yapay zekâ (XAI) teknikleri sayesinde kullanıcılar, modelin neden belli bir varlığı tanıdığı ya da tanımadığı hakkında bilgi sahibi olabilmektedir. Bu da sistemlerin güvenilirliğini artırmakta ve özellikle kamu, sağlık ve hukuk gibi hassas alanlarda kullanımı daha mümkün kılmaktadır. LIME, SHAP ve dikkat ısı haritası gibi XAI yöntemlerinin entegrasyonu bu noktada kritik bir rol oynamaktadır.

Sonuç olarak, Türkçeye özgü dil özelliklerinin dikkate alındığı, alan uyarlaması yapılmış, açıklanabilirliği yüksek ve farklı veri türleriyle test edilmiş NER sistemlerine olan ihtiyaç artmaktadır. Bu çalışma da tam olarak bu ihtiyaç doğrultusunda şekillendirilmiş hem içerik hem yöntem hem de açıklanabilirlik açısından literatüre katkı sağlamayı amaçlamıştır.

3. TEMEL KAVRAMLAR VE YÖNTEMLER

Bu bölümde, çalışma kapsamında kullanılan doğal dil işleme teknikleri, NER sistemleri, derin öğrenme tabanlı dil modelleri, açıklanabilir yapay zekâ yöntemleri ve performans değerlendirme metrikleri ele alınmaktadır. Ayrıca veri seti özellikleri, etiketleme biçimleri ve deneysel ortam hakkında detaylı bilgiler sunulmaktadır.

3.1. Doğal Dil İşleme (Natural Language Processing – NLP)

Doğal dil işleme (NLP), bilgisayarların insan dilini anlaması, yorumlaması ve üretmesi üzerine odaklanan yapay zekâ alanlarından biridir. Bu disiplin; dilbilim, bilgisayar bilimi ve makine öğrenmesi gibi farklı alanların kesişiminde yer almakta ve metin sınıflandırma, özetleme, duygu analizi, bilgi çıkarımı, makine çevirisi gibi geniş bir uygulama yelpazesini kapsamaktadır. Jurafsky ve Martin, NLP'nin temel amacını, yazılı veya sözlü verilerden otomatik olarak anlamlı bilgi elde etmek şeklinde tanımlamaktadır [25].

NLP, dilin çok katmanlı yapısının işlenmesini gerektiren karmaşık bir süreçtir. Bu süreçte; fonetik ve fonoloji, kelimelerin ses yapısını; morfoloji, kök ve ek ilişkilerini; sözdizimi, kelimelerin cümle içindeki dizilim kurallarını; anlambilim, kelime ve cümle düzeyindeki anlam yapılarını; pragmatik ise ifadenin bağlam içindeki kullanım amacını inceler [30]. Bird, Klein ve Loper, NLP'nin farklı uygulama alanlarını detaylandırarak bu teknolojilerin çeşitli sistemlere nasıl entegre edildiğini açıklamışlardır [31]. Örneğin, metin özetleme teknikleri büyük hacimli içeriklerin anlam bütünlüğünü koruyarak kısaltılmasını sağlarken, duygu analizi yöntemleri sosyal medya gönderilerindeki duygusal eğilimleri tespit edebilmektedir. Otomatik çeviri sistemleri çok dilli iletişimi mümkün kılarken, sesli asistanlar gibi uygulamalar kullanıcıların doğal konuşmalarını işleyerek etkileşim sağlamaktadır.

Türkçe için geliştirilen NLP uygulamaları, dilin sondan eklemeli yapısı ve serbest sözdizimi gibi özellikleri nedeniyle ek zorluklar içermektedir. Bir kök sözcüğün çok sayıda türev biçimine sahip olması, kelime ayrıştırma ve anlam çözümleme adımlarını karmaşıklandırmakta; ayrıca kelime dizilişindeki esneklik, bağlam çözümlemesinde daha gelişmiş yöntemlere ihtiyaç doğurmaktadır. Jurafsky ve Martin, bu tür zorlukların aşılmasında derin öğrenme tabanlı, bağlam duyarlılığı yüksek modellerin önemli bir avantaj sağladığını belirtmektedir [25]. Bu nedenle, Türkçeye özgü NLP sistemlerinde morfolojik zenginliği dikkate alan ve bağlamsal temsil gücü yüksek yöntemlerin tercih edilmesi, model

başarımını doğrudan etkilemektedir. Bu tez çalışmasında da mevcut literatürde öne çıkan gelişmeler doğrultusunda, Türkçeye uyarlanmış modeller ve yöntemler esas alınmıştır.

3.2. Dönüştürücü Tabanlı Derin Öğrenme Modelleri

Dönüştürücü mimarisi, doğal dil işleme alanında devrim niteliğinde bir yaklaşım olarak kabul edilmekte olup, birçok dil işleme görevinde yüksek başarı sağlayan derin öğrenme tabanlı modellerin temelini oluşturmaktadır. Vaswani ve arkadaşları tarafından geliştirilen bu yapı, özellikle dikkat mekanizması sayesinde uzun mesafeli bağlam ilişkilerini etkili biçimde modelleyebilme yeteneğine sahiptir [32]. Geleneksel sıralı mimariler yerine tamamen dikkat temelli bir yapıyı benimseyen dönüştürücü mimarisi, paralel işlemeye imkân tanınması sayesinde hem eğitim sürecini hızlandırmakta hem de büyük veri kümeleriyle çalışırken daha verimli sonuçlar elde edilmesini sağlamaktadır.

Dönüştürücü tabanlı modellerin en belirgin avantajlarından biri, çift yönlü bağlam modelleme yetenekleridir. Bu sayede her bir kelimenin anlamı, yalnızca kendisiyle sınırlı kalmadan, içinde bulunduğu cümledeki diğer kelimelerle olan ilişkisi üzerinden değerlendirilmektedir. Özellikle sondan eklemeli yapıya sahip Türkçede, kelime kökleri ile ekler arasında anlam açısından güçlü bağlar bulunduğundan, bu tür bağlam duyarlı öğrenme teknikleri daha isabetli sonuçlar doğurabilmektedir. Bu yönüyle dönüştürücü mimarisi, Türkçe gibi morfolojik açıdan zengin dillerde NER gibi görevler için büyük avantaj sunmaktadır.

Bu tez çalışmasında, Türkçe magazin haberleri üzerinde yürütülen NER görevine yönelik olarak beş farklı dönüştürücü tabanlı model değerlendirilmiştir. Bu modellerden üçü yalnızca Türkçeye özgü verilerle eğitilmişken, diğer ikisi ise çok dilli yapıları sayesinde Türkçeyi de kapsayan önceden eğitilmiş modellerdir. Bu kapsamda kullanılan modeller şunlardır:

- dbmdz/bert-base-turkish-cased (BERTurk)
- google-bert/bert-base-multilingual-cased (mBERT)
- FacebookAI/xlm-roberta-base (XLM-R)
- dbmdz/distilbert-base-turkish-cased (DistilBERTurk)
- dbmdz/convbert-base-turkish-mc4-uncased (ConvBERTurk)

Söz konusu modellerin her biri Hugging Face platformu üzerinden erişilebilir durumdadır. Bu modellerin ön eğitim veri kümesi, dikkat yapısı, katman sayısı, parametre büyüklüğü ve Türkçeye uyarlanma biçimi açısından farklılık göstermektedir. İlgili alt

başlıklarda her model ayrı ayrı ele alınarak mimari özellikleri, dil işleme kapasitesi ve bu tez kapsamında gösterdiği performans detaylı şekilde değerlendirilecektir.

Ayrıca, bu modellere ilişkin mimari yapı, ön eğitim veri kümesi, dikkat mekanizması, katman sayısı, parametre büyüklüğü ve Türkçeye özgünlük düzeyi gibi temel teknik özellikler, karşılaştırmalı olarak Ek 2'de tablo hâlinde sunulmuştur.

3.2.1. BERTurk

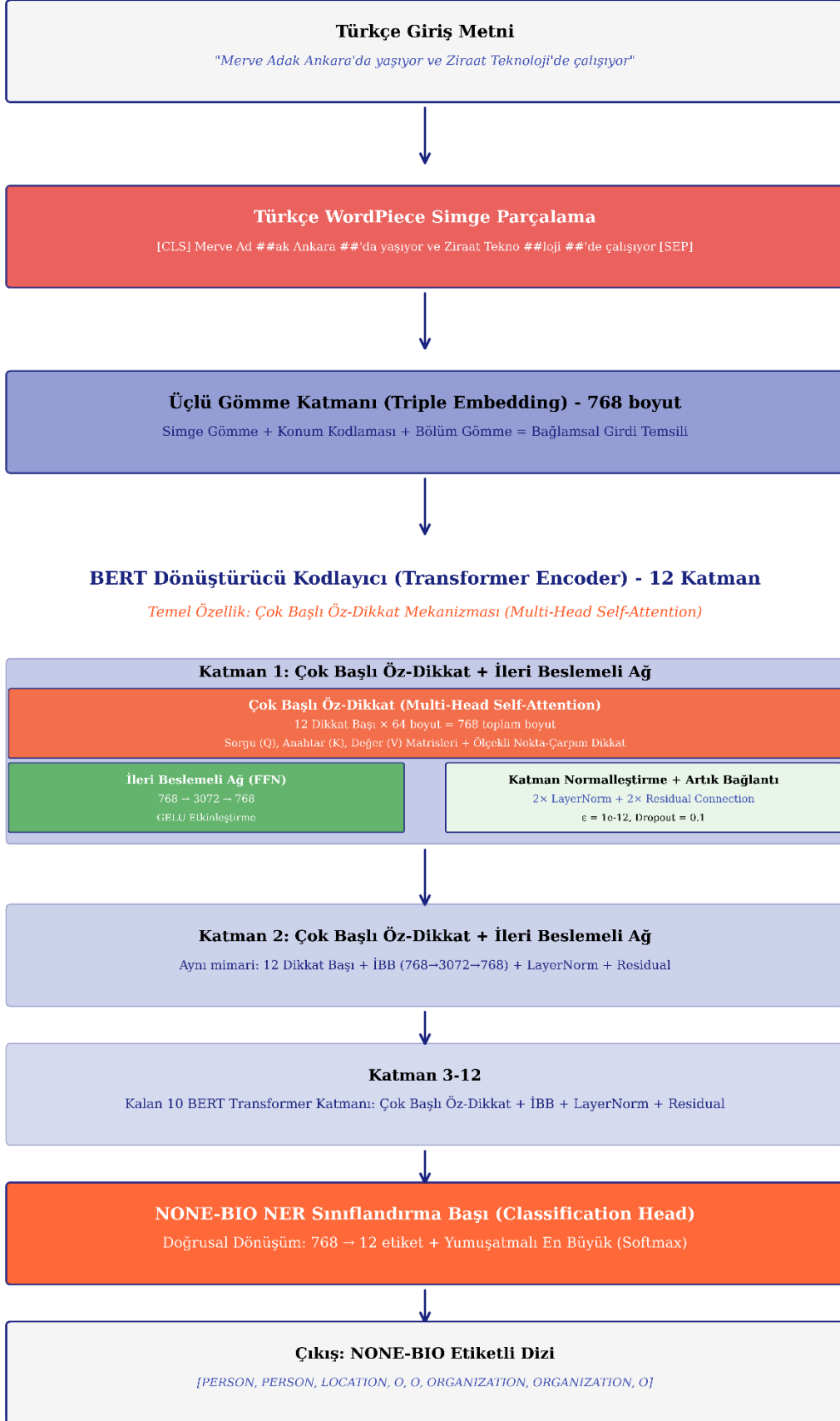
BERTurk modeli, BERT mimarisine dayalı olarak özellikle Türkçe NLP görevleri için geliştirilmiş önceden eğitilmiş bir dil modelidir. Almanya merkezli Bavyera Devlet Kütüphanesi (Bayerische Staatsbibliothek) bünyesinde faaliyet gösteren DBMDZ (Digitale Bibliothek) ekibi tarafından geliştirilen bu model, Türkçeye özel büyük-küçük harf duyarlı bir versiyon olarak sunulmuştur. Model, orijinal BERT mimarisinin çift yönlü bağlam anlayışına dayalı yapısı sayesinde, Türkçenin morfolojik ve sözdizimsel zenginliğini dikkate alarak etkili bir dil temsili sunmayı amaçlamaktadır.

Modelin ön eğitim süreci, yaklaşık 35 GB boyutundaki Türkçe Wikipedia, filtrelenmiş ve cümle bazında parçalara ayrılmış OSCAR korpusu, çeşitli OPUS derlemeleri ve Kemal Oflazer tarafından sağlanan özel metin koleksiyonlarından oluşturulan veri kümesi ile gerçekleştirilmiştir. Toplamda 4,4 milyar birim içeren bu veri, Google'ın TensorFlow Research Cloud altyapısı kullanılarak TPU v3-8 ortamında 2 milyon adım boyunca eğitilmiştir [33].

BERTurk mimarisi, Devlin ve arkadaşları tarafından önerilen orijinal BERT yapısıyla büyük ölçüde paralellik göstermektedir [7]. Modelde 12 adet dönüştürücü katman yer almaktadır. Her bir katman, 12 başlı çoklu dikkat mekanizması ve ileri beslemeli sinir ağı yapısı içermektedir. Her katmanda 768 boyutlu gizli temsiller üretilmekte ve 3072 boyutlu ileri besleme ağı ile işlenmektedir. Giriş metni, 32.000 birimlik WordPiece algoritması ile alt birimlere ayrılmaktadır. Her bir birim için simge, konum ve segment gömmeleri kullanılarak toplam 768 boyutlu bağlamsal temsil vektörü oluşturulmaktadır. Bu temsiller, katman normu ve artık bağlantılar eşliğinde her dönüştürücü model katmanında işlenerek bağlam duyarlılığı yüksek çıktılar elde edilmektedir [34], [35].

Modelin genel işleyişini ve katmanlı mimarisini kavramsal düzeyde görselleştiren yapı Şekil 3.1'de sunulmaktadır. Söz konusu şekil, non-BIO etiketleme formatına uygun olarak örnek bir senaryo üzerinden modellenmiş olup mimarinin bileşenleri açık biçimde görselleştirilmiştir. Giriş aşamasında yer alan WordPiece parçalama süreciyle birlikte, her

alt birim için oluşturulan simge, konum ve segment gömmelerinin birleştirilerek bağlamsal temsillerin oluşturulması gösterilmektedir. Ardından bu temsillerin 12 adet çok başlı dikkat katmanında işlenerek ileri beslemeli ağlara aktarılması ve son aşamada sınıflandırma başlığına ulaşan doğrusal dönüşüm süreci ayrıntılı biçimde betimlenmektedir. Şekil 3.1'deki yapı sayesinde, modelin Türkçe metinler üzerinde nasıl bağlamsal temsil oluşturduğu ve bu temsilleri NER görevinde nasıl kullandığı teknik olarak anlaşılır hale gelmektedir.



Şekil 3.1. BERTurk Model Mimarisi

Bu bütünsel mimari yapı, BERTurk modelini Türkçeye özgü bağlamsal anlam çıkarımında etkili kılmaktadır. Türkçe NER uygulamaları için yüksek doğrulukta ve güvenilir sonuçlar sunan bir temel model olarak öne çıkarmaktadır.

3.2.2. mBERT

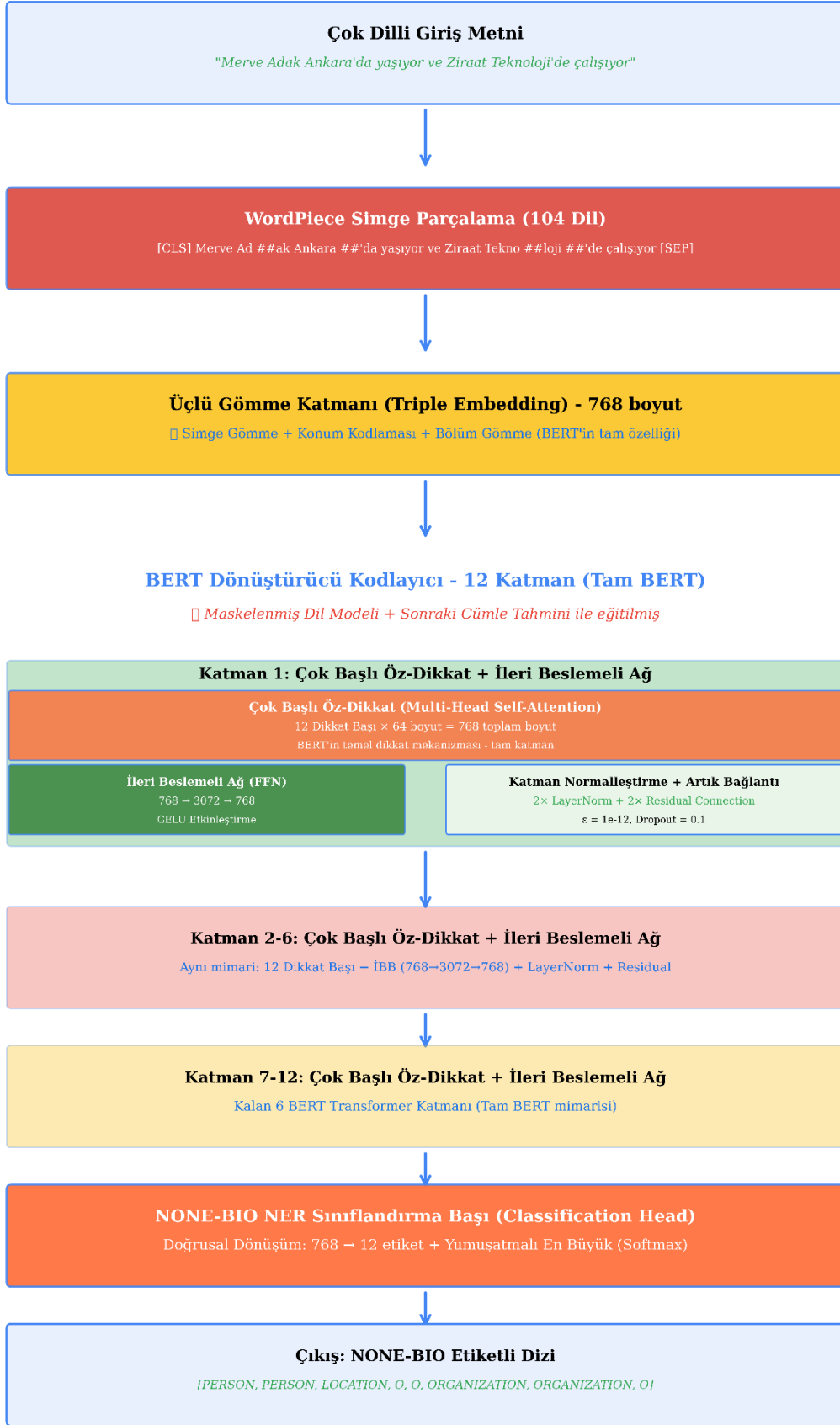
Google tarafından geliştirilen BERT modeli, Devlin ve arkadaşları tarafından 2018 yılında önerilen ve NLP alanında devrim niteliğinde bir yaklaşım sunan çift yönlü kodlayıcı modeldir. Bu mimari, dönüştürücü yapıları temel alarak hem sağdan sola hem de soldan sağa bağlam bilgisini eş zamanlı olarak modelleyebilme yeteneğine sahiptir [7]. Bu doğrultuda geliştirilen Google BERT-Base Multilingual Cased (mBERT) modeli, 104 farklı dildeki Wikipedia verileri üzerinde önceden eğitilmiş çok dilli bir versiyon olup, dil etiketleme, sınıflandırma ve NER gibi görevlerde etkin biçimde kullanılmaktadır [7], [36].

mBERT modeli, 179 milyon parametreye sahiptir. Ön eğitim sürecinde yalnızca ham metinler kullanılmıştır. Herhangi bir insan tarafından etiketleme yapılmamış ve girdiler otomatik olarak maskelenerek öğrenme gerçekleştirilmiştir. Modelin ön eğitimi sırasında iki temel görev kullanılmıştır: Maskelenmiş Dil Modeli (Masked Language Modeling - MLM) ve Sonraki Cümle Tahmini (Next Sentence Prediction - NSP). MLM görevinde, her giriş cümlesindeki simgelerin yüzde 15'i rastgele maskelenmekte ve modelden bu eksik parçaların tahmini istenmektedir. NSP görevinde ise iki cümlenin ardışık olup olmadığı tahmin edilmektedir. Bu iki görev, modelin hem içsel bağlamı hem de cümleler arası ilişkileri öğrenmesini sağlamaktadır.

Modelin girişinde WordPiece simgeleme algoritması kullanılarak cümleler, 110.000 birimlik çok dilli bir sözcük dağarcığına göre alt birimlere ayrılmaktadır. Her bir alt birim için simge, konum ve bölüm gömmeleri ayrı ayrı hesaplanmaktadır. Bu üçlü birleşim sonucunda her simge için 768 boyutlu bağlamsal vektör elde edilmektedir. mBERT mimarisi, 12 adet dönüştürücü katmanı içermekte olup, her bir katman çok başlı öz-dikkat mekanizmasına ve ileri beslemeli sinir ağına sahiptir. Her dönüştürücü katmanında 12 dikkat başı bulunmakta ($12 \times 64 = 768$) ve ardından $768 \rightarrow 3072 \rightarrow 768$ boyutlarında doğrusal dönüşüm katmanları, GELU (Gauss Hatalı Doğrusal Birimi) aktivasyon fonksiyonu, katman normalleştirme ve artık bağlantılar uygulanmaktadır. Maksimum giriş uzunluğu 512 simge ile sınırlandırılmıştır [7], [36], [37].

Modelin genel işlem yapısını, katmanlı mimarisini ve sınıflandırma başlığına giden akışını gösteren kavramsal mimari şema Şekil 3.2'de sunulmaktadır. Şekil 3.2'de WordPiece

simgeleme sürecinden başlayarak üçlü gömme yapısının oluşturulması, her dönüştürücü katmanında gerçekleşen dikkat hesaplamaları ve ileri beslemeli sinir ağı yapısı ile çıkış katmanındaki sınıflandırma başlığının işleyişi görsel olarak detaylandırılmaktadır. Özellikle non-BIO etiketleme formatı dikkate alınarak yapılandırılmış olan bu mimari şema, modelin bilgi işleme sırasını sistematik biçimde ortaya koymakta ve yapıdaki temel teknik bileşenleri tanıtmaktadır. Görselde, 12 katmanlı kodlayıcı yapı, 768 boyutlu temsil vektörleri, 3072 boyutlu ileri beslemeli ağ yapısı, 110.000 sözcüklük çok dilli dağarcık, maksimum 512 birimlik giriş uzunluğu, 0,15 maskelenmiş simge oranı ve çift yönlü bağlam modelleme gibi temel teknik unsurlar öne çıkarılmaktadır.



Şekil 3.2. mBERT Model Mimarisi

3.2.3. XLM-R

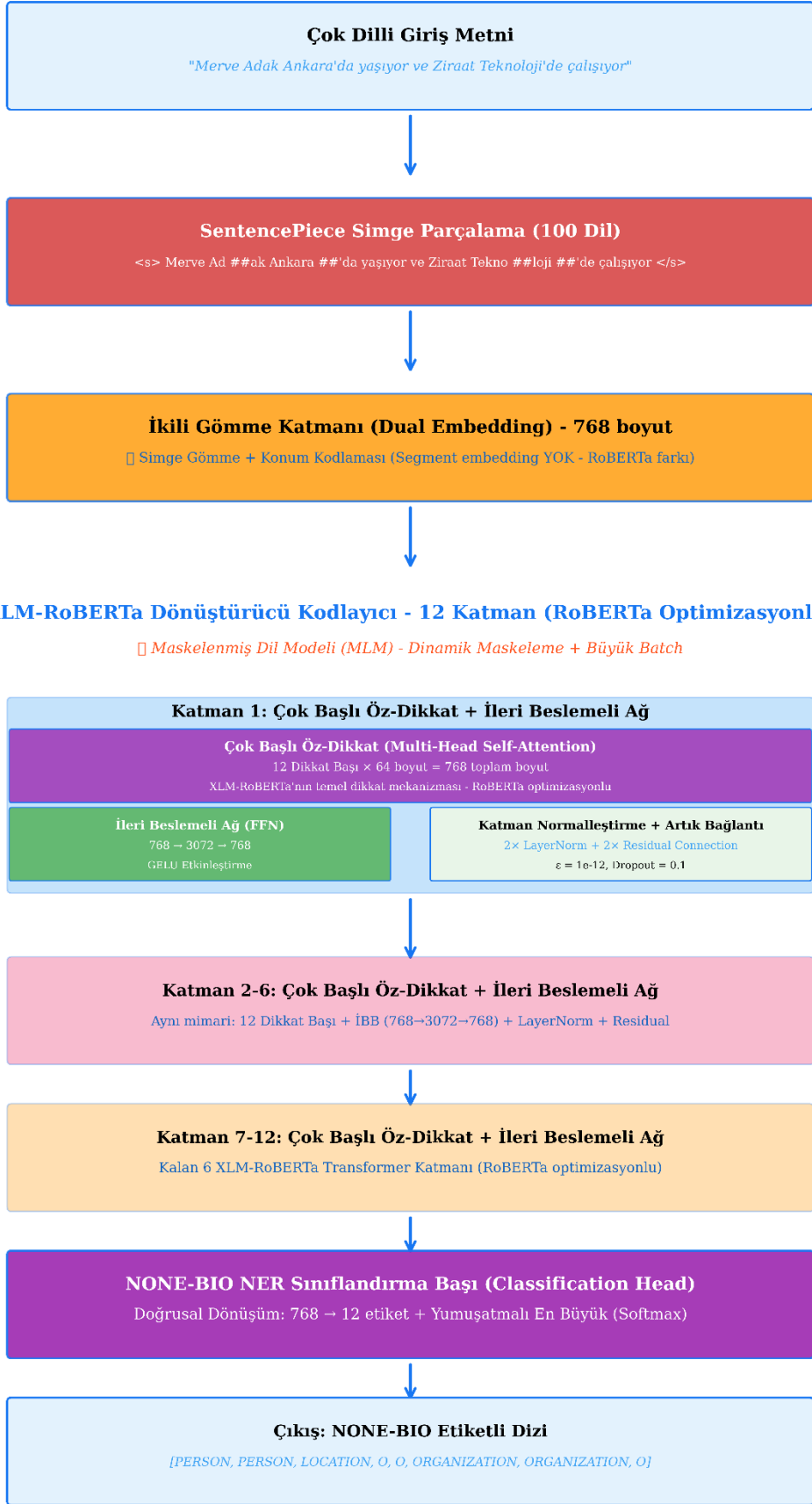
XLM-RoBERTa (Cross-lingual Language Model – Robustly Optimized BERT Pretraining Approach, XLM-R), Facebook AI Research tarafından geliştirilen ve Conneau ile arkadaşları tarafından önerilen, çok dilli bir RoBERTa modelidir. Bu yapı, dönüştürücü temelli çift yönlü kodlayıcı mimari ile RoBERTa'nın optimizasyonlarını bir araya getirerek 100 farklı dilde etkili dil modelleme sağlamaktadır [38].

Bu tez çalışmasında kullanılan XLM-R modeli, CommonCrawl ve Wikipedia veri kümelerinden elde edilen toplam 2.5 TB'lık metin üzerinde yalnızca MLM göreviyle önceden eğitilmiştir. Modelin ön eğitiminde 0,50 oranında dinamik birim maskeleme, büyük yığılı eğitim ve cümle sıralama görevinden feragat edilerek bağlamsal temsillerin daha isabetli öğrenilmesi amaçlanmıştır [39].

XLM-R modelinin simgeleştirme sürecinde, 250.000 birimlik çok dilli bir sözcük dağarcığı kullanan SentencePiece algoritması tercih edilmiştir. Modelin gömme katmanında yalnızca simge ve konum gömmeleri yer almakta olup, BERT yapısından farklı olarak segment gömmesi içermemektedir. Bu yapı, bağlamsal temsil oluştururken gereksiz karmaşıklıkların önüne geçerek daha yalın ve etkili bir gömme stratejisi sunmaktadır [40].

Modelin dönüştürücü katman yapısı, 12 adet kodlayıcı bloktan oluşmaktadır. Her bir blokta 12 başlı öz-dikkat mekanizması, 3072 boyutlu ileri beslemeli sinir ağı, GELU aktivasyon fonksiyonu, katman normu ve artık bağlantılar yer almaktadır. Her dikkat başı 64 boyutunda olup toplamda 768 boyutlu bağlamsal temsiller üretilmektedir. Giriş uzunluğu en fazla 512 simgeyle sınırlandırılmıştır. Toplam parametre sayısı 125 milyon olan model, RoBERTa'nın dinamik maskeleme ve büyük ölçekli eğitim gibi optimizasyonlarını kullanarak daha verimli öğrenme sağlamaktadır [41].

Modelin genel işlem yapısını ve sınıflandırma başlığına kadar olan akışını görselleştiren mimari şema Şekil 3.3'te sunulmaktadır. Bu şema da çok dilli giriş metninden başlayarak SentencePiece simgeleme, ikili gömme katmanı, 12 katmanlı dönüştürücü kodlayıcı yapı ve non-BIO etiketleme formatına uygun sınıflandırma başlığını içeren tam modeli kavramsal olarak sunmaktadır. Görselde, dikkat mekanizmasının işleyişi, ileri beslemeli ağ bileşenleri, katman normları, bağlamsal temsil boyutları, maskeleme stratejisi ve çıkış dizisine dair yapılar açık şekilde gösterilmektedir.



Şekil 3.3. XLM-R Model Mimarisi

XLM-RoBERTa'nın BERT tabanlı modellere göre başlıca farklılıkları, daha gelişmiş bir dil modelleme yeteneği sunması açısından dikkat çekicidir. Öncelikle, segment gömmeleri bu modelde tamamen kaldırılmıştır. Yalnızca simge ve konum gömmeleri kullanılmaktadır. Bu sayede model, daha yalın bir yapı üzerinden bağlamsal bilgi üretmektedir. Ayrıca sabit oranlı maskeleye yerine dinamik birim maskeleye kullanılarak, eğitim sırasında çeşitliliği yüksek ve daha zengin bağlam örnekleri sağlanmaktadır. NSP görevi ise kaldırılarak, yalnızca maskelenmiş dil modeli üzerinden öğrenme gerçekleştirilmiştir. Bu sayede, cümleler arası ilişki zorunluluğu ortadan kaldırılmış ve tekil cümle temsillerine odaklanılmıştır. Simgeleme aşamasında ise, BERT'in WordPiece yerine, daha büyük ve kapsayıcı olan 250.000 birimlik SentencePiece algoritması tercih edilmiştir. Son olarak, modelin ön eğitimi 2.5 TB gibi oldukça geniş bir veri kümesi ile gerçekleştirilmiştir. Bu durum da dilin yapısal çeşitliliğine karşı daha genel geçer temsiller üretilmesine olanak sağlamıştır [38], [39], [40].

Bu güçlü optimizasyon stratejileri sayesinde XLM-RoBERTa, çok dilli bağlamda yüksek performans sergileyen bir yapı ortaya koymakta ve Türkçe gibi morfolojik olarak zengin dillerde bağlam duyarlılığı yüksek çıktılar üretmektedir.

3.2.4. DistilBERTurk

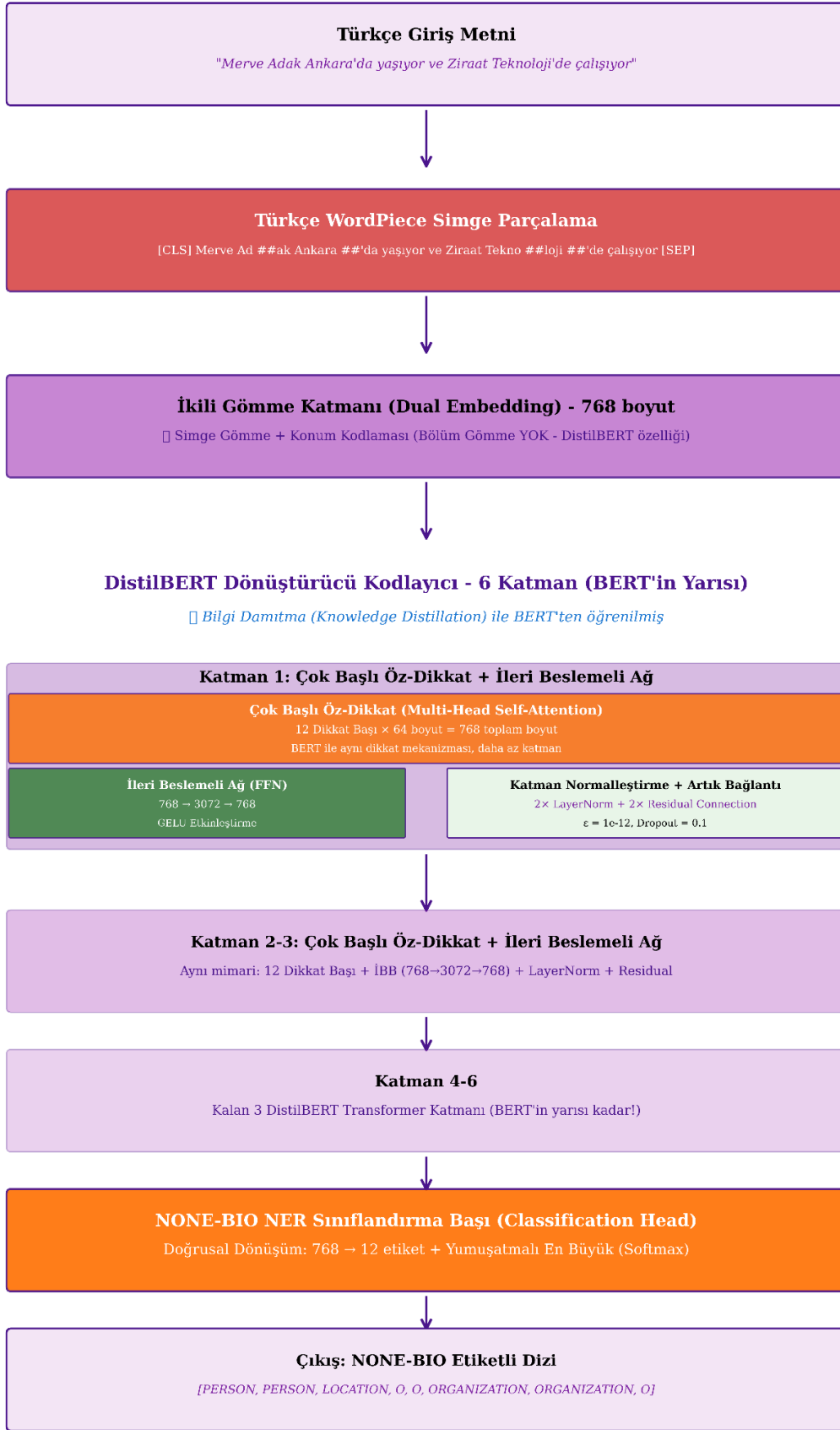
DistilBERTurk modeli, Sanh ve arkadaşları (2019) tarafından önerilen bilgi damıtma (knowledge distillation) tekniği temel alınarak, DBMDZ araştırma ekibi tarafından Türkçeye özel olarak geliştirilmiştir [42]. Bu yapı, öğretmen-öğrenci paradigmasına dayalı olarak BERTurk modelinden türetilmiş olup, daha az sayıda parametre ve daha hızlı çıkarım süresiyle yüksek doğruluk oranını korumayı amaçlamaktadır [43].

DistilBERTurk modelinin bilgi damıtma süreci, üçlü kayıp fonksiyonu ile gerçekleştirilmektedir. Damıtma kaybı (distillation loss) ile öğretmen modelin çıkış dağılımlarının taklidi, kosinüs benzerliği kaybı ile gizli katman temsillerinin korunması, MLM kaybı ile birim düzeyinde öğrenmenin sağlanmasıdır [44], [45]. Bu çok hedefli öğrenme yaklaşımı, modelin hem daha kompakt hale gelmesini hem de öğretmen modelin performansını mümkün olduğunca korumasını sağlamaktadır.

DistilBERTurk modeli, Türkçeye özgü 32.000 birimlik sözcük dağarcığı ile WordPiece simgeleme algoritmasını kullanır. BERTurk ile aynı yaklaşımını benimseyen model, simge ve konum gömmelerini kullanarak ikili bir gömme stratejisi izlemektedir. Her giriş simgesi, 768 boyutlu vektörlerle temsil edilir.

Modelin mimarisi, 6 adet dönüştürücü kodlayıcı katmandan oluşur ve her bir katmanda 12 başlı çoklu dikkat mekanizması ($12 \times 64 = 768$), $3072 \rightarrow 768$ boyutlarında ileri beslemeli sinir ağı, GELU aktivasyon fonksiyonu, katman normalleştirme ($\epsilon = 1e-12$) ve artık bağlantılar yer almaktadır [46]. Bu yapı, orijinal BERT'in yarısı kadar katman içerirken, dikkat mekanizmaları ve çıktı boyutları açısından öğretmen modelle aynı yapıyı korur. Model toplamda yaklaşık 66 milyon parametre içerir ve 0,60 daha küçük olmasına rağmen BERTurk modelinin performansının 0,97'sini koruyacak biçimde optimize edilmiştir.

Modelin genel işlem yapısını ve non-BIO etiketleme formatına göre sınıflandırma başlığına kadar olan mimarisini kavramsal olarak sunan yapı Şekil 3.4'te gösterilmektedir. Bu şema da Türkçe giriş metninin WordPiece simgelere ayrılmasından başlayarak ikili gömme katmanı, altı adet dönüştürücü blok, her bir blokta yer alan dikkat mekanizmaları, ileri beslemeli ağ bileşenleri, katman normu ve artık bağlantılar ile son aşamada yer alan sınıflandırma başlığına kadar tüm süreci ayrıntılı şekilde görselleştirmektedir. Şemada, segment gömmesinin çıkarıldığı, yalnızca simge ve konum gömmeleriyle çalışıldığı, her katmanda çok başlı dikkat yapısının korunduğu ve final katmanda 768 boyutlu çıktının 12 sınıfa indirgenerek yumuşatmalı en büyük aktivasyon fonksiyonu ile tahminlerin üretildiği açık biçimde gösterilmektedir.



Şekil 3.4. DistilBERTurk Model Mimarisi

3.2.5. ConvBERTurk

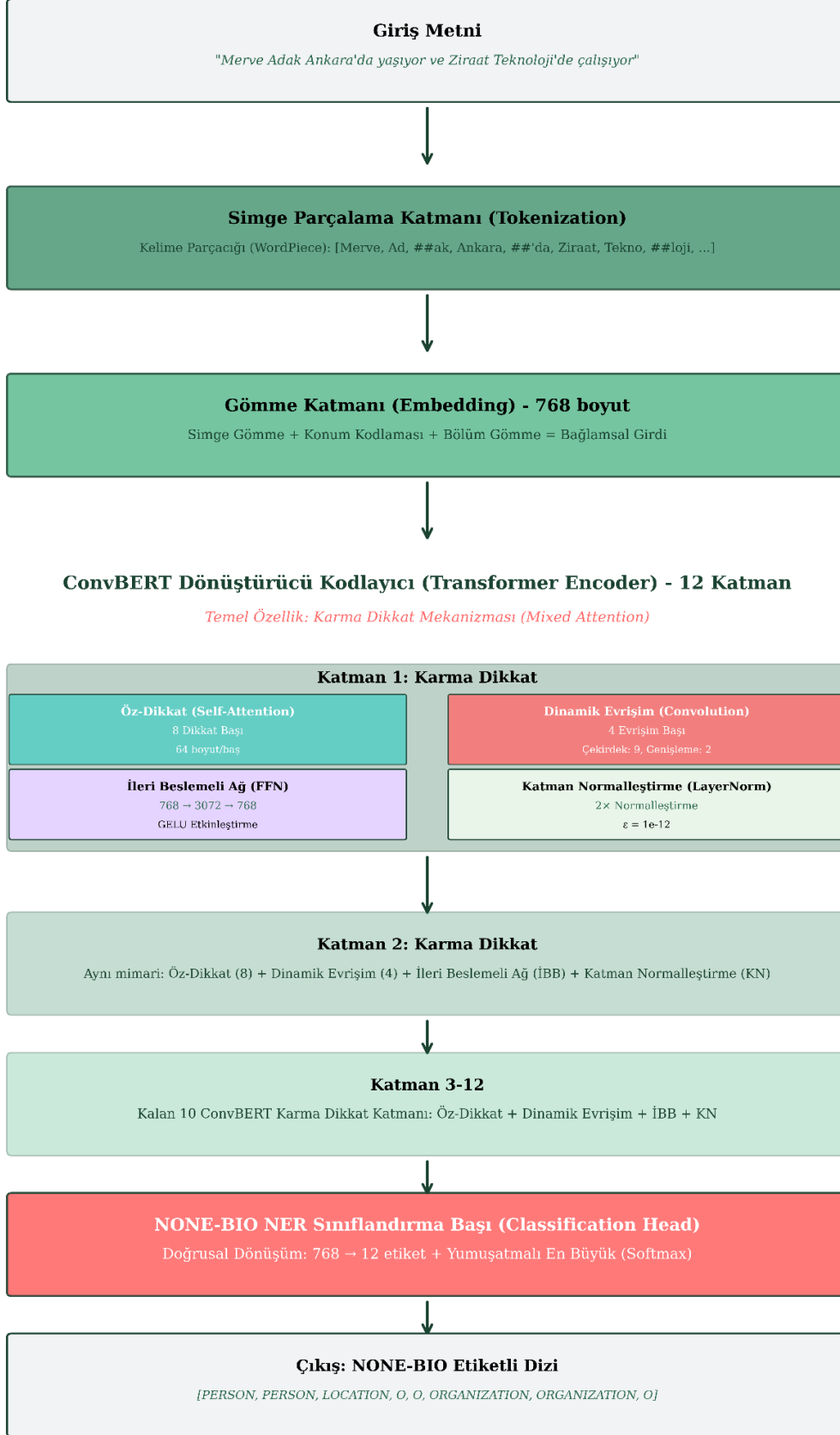
ConvBERTurk modeli, Jiang ve arkadaşları (2020) tarafından geliştirilen ConvBERT mimarisi temel alınarak, DBMDZ araştırma grubu tarafından Türkçeye özgü olarak eğitilmiştir [47]. ConvBERT, geleneksel BERT mimarisine dinamik evrişim katmanlarını entegre ederek karma dikkat mekanizmasını benimseyen bir yapıdır. Bu mimari sayesinde hem global hem de yerel bağımlılıklar eş zamanlı olarak modellenenmektedir.

Türkçe için geliştirilen ConvBERTurk modeli, AllenAI tarafından yayımlanan 242 GB büyüklüğündeki Türkçe Multilingual C4 veri kümesi üzerinde eğitilmiştir. Modelde orijinal 32.000 sözcük dağarcığı kullanılmış, eğitim süreci 1 milyon adım boyunca 512 uzunlukta dizilerle ve TPU (v3-32) altyapısıyla gerçekleştirilmiştir [48].

ConvBERTurk modeli, WordPiece simgeleme algoritması ile giriş metinlerini alt parçalara ayırır ve bu parçaları simge, konum ve segment gömmeleri ile 768 boyutlu vektörler şeklinde temsil eder. Toplamda 12 kodlayıcı katman içeren modelde her katmanda karma dikkat mekanizması yer almakta olup, bu mekanizma iki ana bileşenden oluşur. Bunlar 8 başlı öz-dikkat ve 4 başlı dinamik evrişim modülleridir. Her modül sonrası katman normalleştirme ($\epsilon = 1e-12$) uygulanır, ardından $768 \rightarrow 3072 \rightarrow 768$ boyutlarında iki doğrusal dönüşüm içeren ileri beslemeli sinir ağı yer almaktadır.

Modelin parametre sayısı yaklaşık 110 milyondur. NER görevi için, modelin son katmanına doğrusal bir sınıflandırıcı eklenmiş olup non-BIO etiketleme sistemine uygun şekilde 12 etiketli bir çıktı üretilmektedir. Çıktı vektörü yumuşatmalı en büyük (Softmax) aktivasyon fonksiyonu ile normalize edilmektedir.

Modelin genel işlem yapısı ve non-BIO etiketleme başlığına kadar olan mimarisi Şekil 3.5'te kavramsal olarak sunulmaktadır. Şekilde, giriş metninin WordPiece simgelerine bölünmesi, gömme katmanları (simge + konum + segment), 12 katmanlık dönüştürücü mimarisi içerisinde öz-dikkat ve dinamik evrişim birimlerinin yerleşimi ve son sınıflandırma adımı ayrıntılı olarak gösterilmektedir.



Şekil 3.5. ConvBERTurk Model Mimarisi

3.3. Açıklanabilir Yapay Zekâ (Explainable AI – XAI)

Gelişen derin öğrenme algoritmaları yüksek doğruluk oranlarına ulaşsa da bu modellerin karar verme süreçleri çoğunlukla kullanıcılar için anlaşılması güç kara kutu yapılar hâlini almıştır. Bu durum, özellikle güvenilirlik ve etik sorumluluk açısından açıklanabilirlik ihtiyacını gündeme getirmiştir [49].

Doğal dil işleme bağlamında XAI yöntemleri, modellerin hangi kelimelere, yapısal ilişkilere veya bağlamsal ipuçlarına dayanarak karar verdiğini analiz etmeyi mümkün kılar. Böylece yalnızca modelin çıktısı değil, bu çıktının nasıl üretildiği de görünür hâle gelir. Özellikle NER gibi çok katmanlı görevlerde, açıklanabilirlik analizleri modelin bağlam duyarlılığı ve sınıflandırma davranışları hakkında derinlemesine içgörüler sunar.

Bu çalışmada açıklanabilirlik amacıyla kullanılan yöntemlerden biri, Ribeiro, Singh ve Guestrin tarafından geliştirilen LIME (Local Interpretable Model-Agnostic Explanations) tekniğidir. LIME, herhangi bir tahminleme modelinin belirli bir örnek üzerinde nasıl bir karar verdiğini anlamaya yönelik yerel doğrusal modellemeler oluşturur. Modelin sınıflandırdığı örneğe benzer sentetik veriler üreterek, bu küçük veri kümesi üzerinde yorumlanabilir ve basit bir model eğitir. Böylece karmaşık yapılar, lokal düzeyde anlaşılabilir hâle getirilir [50].

Modelden bağımsız yapısıyla LIME, derin öğrenme tabanlı sistemlerin karar mekanizmalarını şeffaflaştırmada etkili bir araçtır. NER sistemlerine uygulandığında, bir kelimenin neden belirli bir etiketle işaretlendiği kelime bazlı katkı puanlarıyla ortaya konabilir. Pozitif ya da negatif etkiye sahip sözcüklerin katkı düzeyleri renk kodlu haritalar aracılığıyla görselleştirilerek yorumlanabilir hâle gelir.

Bu tezde LIME yöntemi; NER modellerinin karar süreçlerini şeffaflaştırmak, bağlam duyarlılıklarını değerlendirmek ve sınıflandırma hatalarının nedenlerini açıklamak amacıyla kullanılmıştır.

3.4. Değerlendirme Metrikleri

NER modellerinin doğruluğunu değerlendirmek amacıyla bu tez kapsamında çeşitli sayısal ölçütler kullanılmıştır. Bu metrikler, modellerin genel başarımını ve sınıf bazlı güçlü/zayıf yönlerini analiz etmeye olanak tanımaktadır [2], [51]. Kullanılan metriklerin tanımları ve işlevleri izleyen alt başlıklarda ayrıntılı biçimde ele alınmıştır.

3.4.1. Kesinlik, Duyarlılık ve F1 Skoru

Kesinlik (precision), modelin olumlu sınıfa ait olarak işaretlediği örneklerin ne kadarının gerçekten doğru olduğunu; duyarlılık (recall), gerçek olumlu örneklerin ne kadarının doğru tahmin edildiğini ifade eder. F1 skoru ise bu iki değer harmonik ortalamasıdır ve özellikle dengesiz sınıf dağılımlarında model başarımını özetleyen en yaygın ölçütlerden biridir [51], [52].

3.4.2. Makro ve Ağırlıklı Değerlendirme Metrikleri

Sınıflar arasında dengesizlik içeren NER problemlerinde, klasik metrikler tek başına yeterli olmamaktadır. Bu nedenle, makro ve ağırlıklı değerlendirme metrikleri tercih edilmiştir. Makro F1 skoru, her sınıf için hesaplanan F1 skorlarının aritmetik ortalamasını alarak tüm sınıflara eşit ağırlık verir. Ağırlıklı F1 ise her sınıfın veri kümesindeki temsil oranını dikkate alarak ağırlıklı ortalama hesaplar. Böylece dengesiz veri kümelerinde daha gerçekçi bir değerlendirme sağlanır [53].

3.4.3. Karışıklık Matrisi

Karışıklık matrisi (confusion matrix), modelin hangi sınıflarda doğru ve hatalı tahminler yaptığını görselleştirerek hata analizi yapmaya olanak tanır. Özellikle sık karıştırılan sınıfların (örneğin "B-LOCATION" ve "B-ORGANIZATION") tespiti ve sınıf içi/genel hata yapısının analizi açısından önemli bir araçtır [54].

3.4.4. ROC-AUC ve PR Eğrileri

ROC eğrisi, farklı eşik değerlerinde modelin doğru pozitif ve yanlış pozitif oranlarını görselleştirmektedir. PR eğrisi ise özellikle dengesiz veri kümelerinde daha anlamlı sonuçlar sunmaktadır [55].

3.4.5. t-SNE ile Görselleştirme

Yüksek boyutlu vektör temsillerini iki boyutlu düzleme indirgemeye yarayan t-SNE (t-distributed stochastic neighbor embedding) yöntemi, modelin öğrenme sürecinde oluşturduğu sınıf temsillerinin ayrışma düzeyini görselleştirmek amacıyla kullanılmıştır. Özellikle benzer sınıfların birbirine yakın kümelenip kümelenmediği, sınıf sınırlarının netliği gibi yapılar t-SNE grafikleri ile gözlemlenmiştir [56].

3.4.6. Varlık Seviyesi ve Sözcük Birimi Seviyesi Değerlendirme

NER sistemleri hem sözcük hem de varlık düzeyinde değerlendirilebilmektedir. Sözcük düzeyinde her bir kelimenin doğru etiketlenip etiketlenmediği ölçülürken, varlık düzeyinde bir varlık bütününün başlangıç ve bitiş konumlarının tam doğruluğu esas alınır [2]. Bu tezde kullanılan non-BIO etiketleme formatı yalnızca sözcük düzeyi değerlendirmeye uygundur. Buna karşılık, BIO formatı ile eğitilen modellerde her iki düzeyde analiz gerçekleştirilmiştir.

3.5. Veri Seti

Bu tez çalışmasında kullanılan veriler, Hugging Face platformunda yayımlanan ve Anıl Güven tarafından derlenen Turkish News Dataset [57] temel alınarak oluşturulmuştur. Ancak yalnızca bu hazır içerikle yetinilmeyip, tez kapsamında doğrudan tarafımda sıfırdan oluşturulmuş, NER görevine uygun özgün bir veri kümesi geliştirilmiştir. Başlangıçta referans alınan veri seti; ekonomi, politika, yaşam, teknoloji, magazin, sağlık ve spor olmak üzere yedi farklı kategoride toplam 4.200 Türkçe haber başlığından oluşmaktadır.

Çalışma kapsamında yalnızca magazin kategorisindeki içerikler seçilmiş ve bu içerikler doğrudan tarafımdan genişletilerek yeniden yapılandırılmıştır. Özellikle MagazinBurada.net, Haberler.com ve Milliyet Magazin gibi güncel dijital platformlardan manuel olarak seçilen örnekler doğrultusunda içerik çeşitliliği sağlanmış ve gerçek dünyadaki dil kullanımını yansıtılmaya çalışılmıştır. Bu süreçte gerek cümlelerin derlenmesi gerekse etiketleme işlemi bizzat tarafımdan yürütülmüş olup, her bir örnek özenle hazırlanmıştır. Böylece hazır bir veri setinden türemeyen, tamamen elle oluşturulmuş, denetimli ve NER görevine özel olarak tasarlanmış bir veri kümesi elde edilmiştir.

Oluşturulan bu veri kümesi, NER görevine uygun olarak iki farklı etiketleme şemasına göre düzenlenmiştir: BIO ve non-BIO formatları. Etiketleme sürecinde, magazin içeriklerine özgü olarak PERSON, LOCATION, NUMBER/DATE, ORGANIZATION, MOVIE/TV, SONG/ALBUM, EVENT, SOCIAL_MEDIA, AWARD, RELATIONSHIP ve HEALTH gibi varlık türleri tanımlanmıştır. Her kelime bağlamına göre tarafımda uygun şekilde etiketlenmiştir. Etiketli veri, CoNLL (Conference on Natural Language Learning) yapısına benzer biçimde düzenlenmiştir. Her bir kelime, karşılık gelen etiketiyle hizalanmıştır.

Örnek olarak, aynı cümle için BIO ve non-BIO biçimlerine ait örnek bir etiketleme gösterilmektedir.

Cümle: Aleyna Tilki, Akasya AVM’de görüntülendi.

- BIO etiketleme biçimiyle:

Aleyna	B-PERSON
Tilki,	I-PERSON
Akasya	B-LOCATION
AVM'de	I-LOCATION
görüntülendi.	O
- non-BIO etiketleme biçimiyle:

Aleyna	PERSON
Tilki,	PERSON
Akasya	LOCATION
AVM'de	LOCATION
görüntülendi.	O

BIO formatı, çok kelimeli varlıkları başlama ve iç etiketleriyle ayrıştırarak daha detaylı bir yapıya olanak tanırken; non-BIO formatı, yalnızca varlık türünü belirtmekle yetinerek daha sade bir etiketleme sağlamaktadır.

Her iki formatta da sınıf dağılımları, Ek 3'te detaylı olarak sunulmuştur. Bu dağılımlar, özellikle model performanslarının karşılaştırılmasında, sınıf dengesizliği etkisinin yorumlanmasında ve değerlendirme metriklerinin anlamlı şekilde analiz edilmesinde önemli bir referans niteliği taşımaktadır.

3.6. Deneysel Ortam ve Sistem Gereksinimleri

Tez kapsamında yürütülen model eğitimleri, değerlendirme analizleri ve açıklanabilirlik çalışmaları, Linux tabanlı bir sanal makine ortamında gerçekleştirilmiştir. Kullanılan sanal makine, doğal dil işleme görevleri için yeterli donanımsal kapasite sağlayacak biçimde yapılandırılmış olup 8 sanal işlem birimi (vCPU), 32 GB RAM ve 512 GB SSD disk alanına sahiptir. Mimari yapı x86_64 olup AMD-V sanallaştırma teknolojisini desteklemektedir. Kodlama süreci PyCharm IDE aracılığıyla gerçekleştirilmiş, tüm deneylerde Python 3.12 sürümü kullanılmıştır. Bu donanımsal yapı sayesinde model eğitim süreçleri kesintisiz yürütülmüş, değerlendirme ve görselleştirme işlemleri güvenilir şekilde tamamlanmıştır. Özellikle belleğin yeterliliği ve işlemci kapasitesi, yüksek boyutlu dönüştürücü tabanlı modellerle yapılan işlemlerde sistemsel darboğaz yaşanmasını önlemiştir. Deneysel süreçlerin istikrarlı şekilde ilerlemesini mümkün kılmıştır.

4. DENEY ÇALIŞMALARI

4.1. Deneysel Kurulum ile Eğitim ve Değerlendirme Süreci

Bu tez çalışmasında, Türkçe magazin içeriklerinden türetilmiş NER görevine yönelik olarak iki farklı etiketleme sistemi kullanılmıştır. Bunlar, çok kelimeli varlıkları başlangıç ve iç etiketleriyle belirginleştiren BIO formatı ve daha yalın bir etiketleme yaklaşımı sunan non-BIO formatıdır. Her iki etiketleme biçimi için beş farklı dönüştürücü tabanlı model (BERTurk, mBERT, XLM-RoBERTa, DistilBERTurk ve ConvBERTurk) eğitilmiş ve ayrı ayrı değerlendirilmiştir. Böylece hem veri etiketleme şemasının hem de model altyapısının başarımlar üzerindeki etkisi karşılaştırmalı olarak analiz edilmiştir.

Model eğitim süreci, Hugging Face platformunun sağladığı Trainer arayüzüyle gerçekleştirilmiş ve erken durdurma yöntemi kullanılmıştır. Yani model, en fazla 1000 döngüye (epoch) kadar eğitilecek şekilde ayarlanmış, ancak doğrulama başarımlarını birkaç ardışık denemede artış göstermediği anda eğitim otomatik olarak sonlandırılmıştır. Burada kullanılan eşik değeri 0,001 olarak belirlenmiş, art arda 3 kez gelişme gözlenmemesi hâlinde süreç durdurulmuştur. Bu parametreler, Liu ve arkadaşlarının doğal dil işleme görevlerinde önerdiği yapılandırmalarla örtüşmektedir [41]. Ayrıca Hussein ve Shareef'in yaptığı deneysel çalışmada da bu tür küçük sabır değerlerinin modelin daha erken kararlı sonuçlar üretmesini sağladığı ve aşırı öğrenmeyi engellediği ifade edilmiştir [58].

Tüm modeller için sabit öğrenme oranı olarak $2e-5$ değeri tercih edilmiştir. Bu seçim, Jurafsky ve Martin'in doğal dil işleme modellerine yönelik önerdiği hiper parametre aralıklarıyla örtüşmektedir [25]. Ayrıca deneylerde kullanılan toplu işlem boyutları (batch size) 8, 16 ve 32 olmak üzere çeşitlendirilmiştir. Model yapısında yer alan bırakma oranı (dropout) ise tüm koşullarda sabit olarak 0,1 tutulmuştur.

Eğitim ve değerlendirme süreçleri, 5 katlı çapraz doğrulama yöntemi ile gerçekleştirilmiş ve toplamda 15 farklı eğitim çalışması yürütülmüştür. Bu yöntem sayesinde, modelin farklı veri alt kümeleri üzerindeki başarımlarını gözlemlenmiş ve sonuçların genellenebilirliği artırılmıştır. Her döngü sonunda modelin doğrulama başarımlarını; kesinlik, duyarlılık ve F1 skoru gibi metriklerle değerlendirilmiştir.

Etiketleme sistemlerinin teknik farkları da eğitim sürecinde dikkate alınmıştır. Özellikle BIO formatı kullanıldığında, kelimelerin alt parça sözcük birimlerine bölünmesi nedeniyle ortaya çıkabilecek hizalama problemlerinin önüne geçmek amacıyla, Hugging Face tarafından sağlanan ofset eşleme (offset_mapping) özelliğinden faydalanılmıştır.

Örneğin “Ankaralılar” gibi bir kelimenin alt parçalara ([An, ##kara, ##lı, ##lar]) ayrıldığı durumlarda yalnızca ilk parçaya etiket atanması gerekmektedir. Bu durumda ofset eşleme kullanılarak yalnızca orijinal kelimenin başlangıç konumundaki alt parçaya etiket verilmiş ve geri kalanlar dışlanmıştır. Böylece model tahminleri ile gerçek kelimeler hizalanmış, yanlış bölünmeden kaynaklanabilecek ölçüm hataları engellenmiştir.

Buna karşılık non-BIO formatı, alt parçalara ayrılmadan doğrudan kelime düzeyinde etiketleme yaptığı için hizalama sorunları yaşanmamış ve daha yalın bir süreç izlenmiştir.

Sonuç olarak, eğitim verisinin sabit tohum değeri ile bölünmesi, hiper parametrelerin güncel literatür ışığında dikkatle seçilmesi ve etiketleme sistemlerinin teknik ayrıntılarının titizlikle ele alınması sayesinde modellerin karşılaştırmalı başarımları hem adil hem de tekrarlanabilir biçimde değerlendirilmiştir.

4.2. Model Bazlı Performans Sonuçları

Bu bölümde, beş farklı dönüştürücü tabanlı modelin Türkçe magazin haberleri üzerinde eğitilip test edildiği deneysel sonuçlar sunulmuştur. Değerlendirme hem BIO hem de non-BIO etiketleme şemalarına göre yapılmıştır. Genel metriklerin yanı sıra sınıf bazlı başarılar ve görsel analizlerle desteklenmiştir. Her model için en iyi sonuç alınan konfigürasyon ve katman esas alınarak performanslar ayrıntılı biçimde yorumlanmıştır.

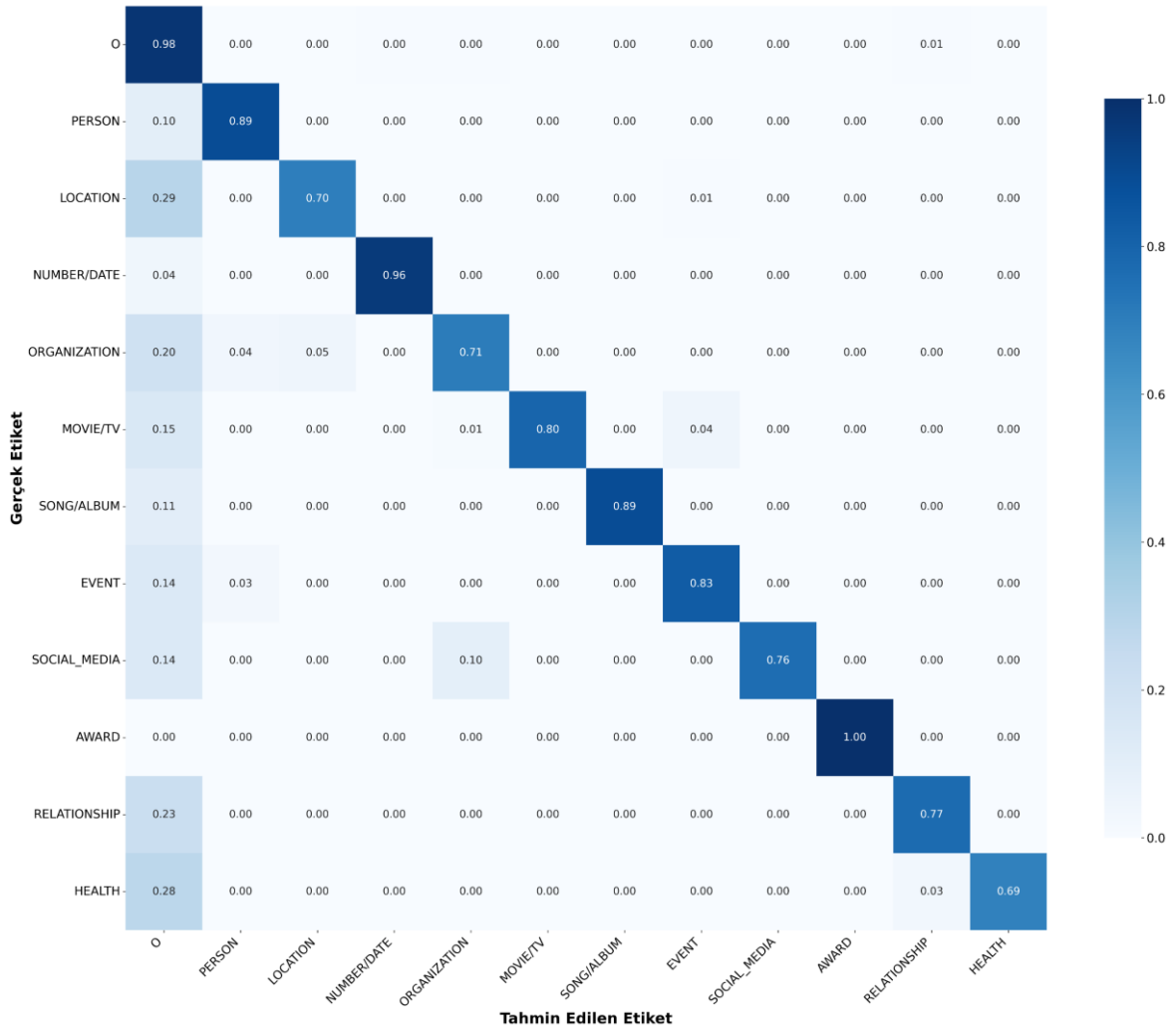
Eğitim sürecinde her model, 5 katlı çapraz doğrulama yöntemiyle test edilmiş ve üç farklı toplu işlem boyutu (batch size: 8, 16 ve 32) denenerek kapsamlı bir hiper parametre araması yapılmıştır. Eğitim döngüsü için azami 1000 sınırı belirlenmiş, ancak tüm katlarda erken durdurma (early stopping) mekanizması devreye girerek erken durdurma sağlanmıştır. Böylece, modelin aşırı öğrenme eğiliminden kaçınılmış, aynı zamanda eğitim süreleri de ciddi oranda kısaltılmıştır.

4.2.1. BERTurk (non-BIO)

BERTurk modeli, non-BIO etiketleme biçiminde gerçekleştirilen deneylerde yüksek doğruluk ve istikrarlı sınıf bazlı performans göstermiştir. Eğitim sürecinde tüm denemelerde erken durdurma mekanizması etkin biçimde çalışmış, ortalama eğitim döngüsü 9,3 ve toplam süre 154,7 dakika olarak kaydedilmiştir. Bu veriler, modelin kısa sürede yüksek başarımlar seviyesine ulaşabildiğini ve veri setine hızlı uyum sağladığını göstermektedir. Farklı toplu işlem boyutları ve veri bölünmeleriyle elde edilen ayrıntılı sonuçlar Ek 4’te sunulmuştur. En

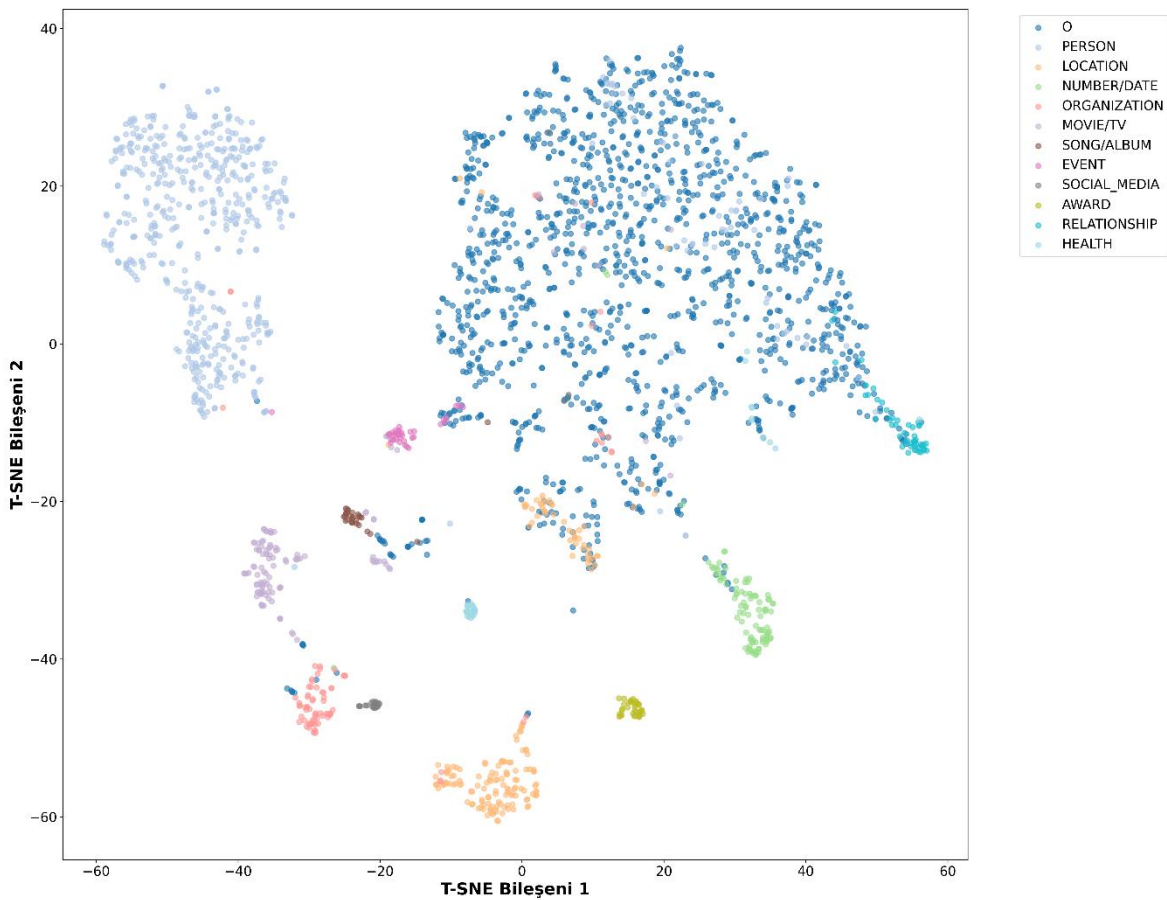
yüksek test makro F1 değeri 0,8728 olup, bu sonuç 16'lık toplu işlem boyutuna sahip birinci katmanda elde edilmiştir.

Modelin sınıf bazlı başarımı, Şekil 4.1'de verilen Normalize Edilmiş Karışıklık Matrisi ile görsel olarak değerlendirilmiştir. Matriste diyagonal üzerinde yer alan yüksek değerler, modelin doğru sınıflandırma oranlarının genel olarak yüksek olduğunu göstermektedir. Özellikle O, NUMBER/DATE, PERSON ve SONG/ALBUM etiketlerinde doğruluk oranlarının 0,89 ile 0,98 arasında değiştiği görülmektedir. Bununla birlikte HEALTH, RELATIONSHIP ve LOCATION etiketlerinde doğru sınıflandırma oranlarının görece daha düşük seviyelerde kaldığı tespit edilmiştir. AWARD etiketinde ise kesinlik oranı 1,00 seviyesinde olup model bu sınıf için hiç yanlış pozitif üretmemiştir; ancak duyarlılık değeri daha düşük olup, bazı doğru örneklerin atlandığı anlaşılmaktadır.



Şekil 4.1. Normaliz Edilmiş Karışıklık Matrisi (BERTurk, non-BIO)

Modelin çıktıları daha detaylı olarak Şekil 4.2’de sunulan t-SNE görselleştirmesi ile değerlendirilmiştir. Bu analiz, modelin ürettiği kelime temsillerini iki boyutlu düzleme indirirken her bir örneği ait olduğu etiketle renklendirerek sunmaktadır. Bu görselleştirmede, PERSON, LOCATION, MOVIE/TV, SONG/ALBUM ve ORGANIZATION gibi sık görülen sınıfların kendi içinde belirgin kümeler oluşturduğu görülmektedir. Buna karşın EVENT, HEALTH, RELATIONSHIP ve SOCIAL_MEDIA gibi sınıfların daha dağınık bir dağılıma sahip olduğu ve bazı örneklerin diğer sınıflarla çakıştığı gözlenmektedir. Bu durum, ilgili sınıflardaki veri çeşitliliğinin ve miktarının artırılması hâlinde modelin ayırım gücünün artabileceği beklenmektedir.



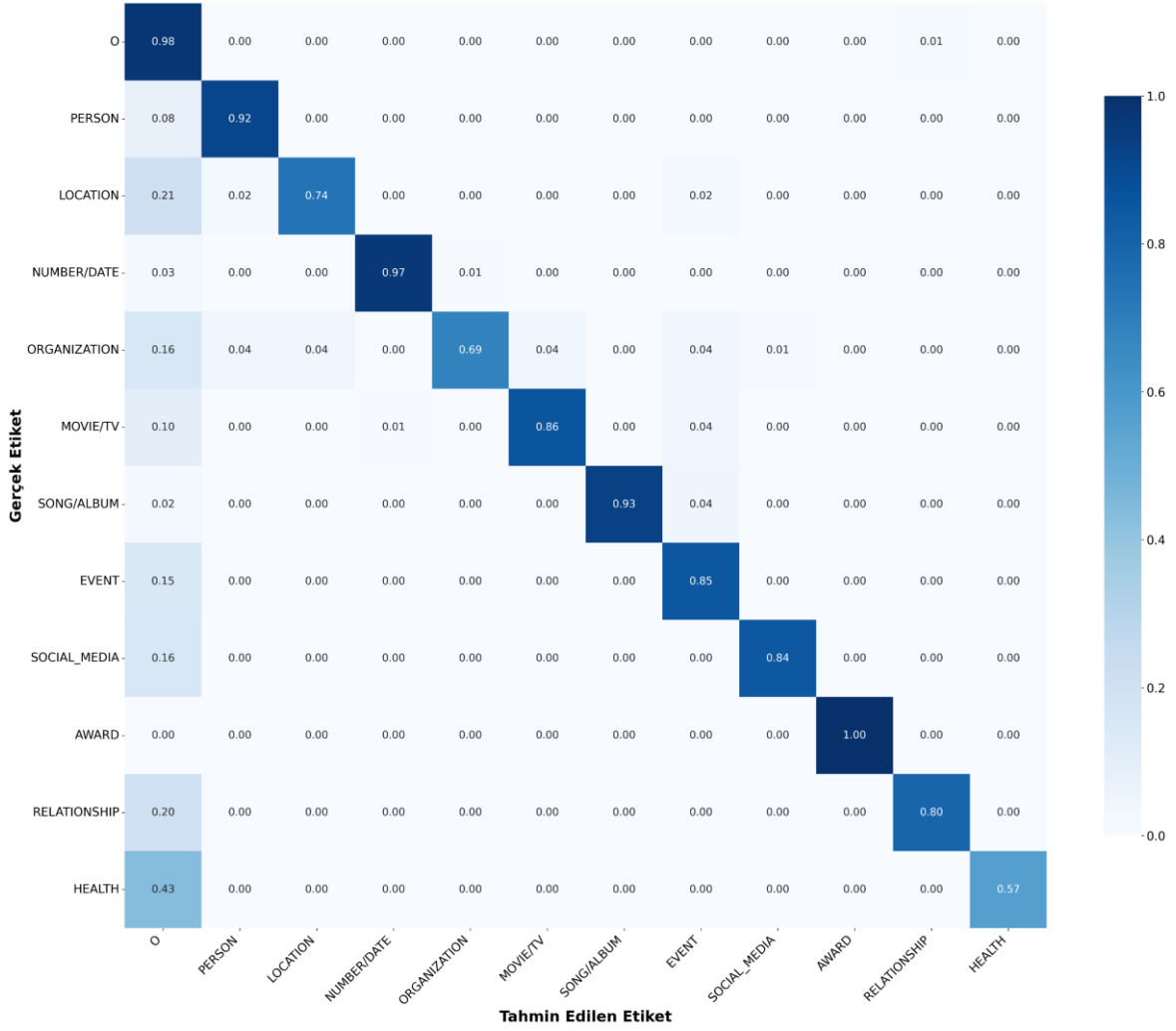
Şekil 4.2. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (BERTurk, non-BIO)

Genel olarak BERTurk modeli, non-BIO formatta güçlü ve dengeli bir performans ortaya koymuştur. Karışıklık matrisi, yaygın etiketlerde yüksek doğruluk yakaladığını; t-SNE ise bu etiketlerin net kümeler hâlinde ayrıştığını göstermektedir. Az görülen veya benzer anlamlı etiketlerde ise zaman zaman karışmalar yaşansa da modelin genel tahminleri güvenilir düzeydedir.

4.2.2. mBERT (non-BIO)

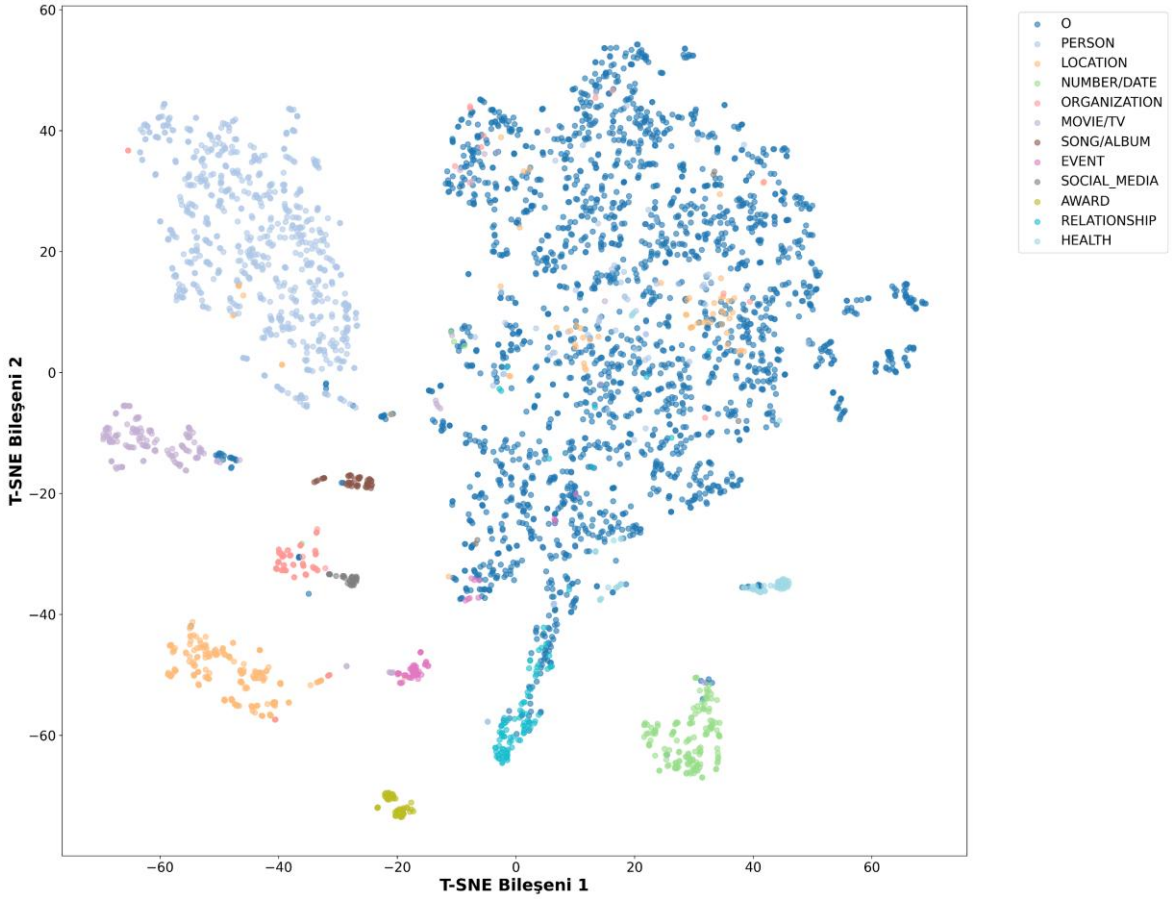
Google tarafından geliştirilen çok dilli BERT modeli olan mBERT, non-BIO etiketleme biçimiyle yapılan deneylerde yüksek doğruluk ve dengeli sınıf bazlı performans sergilemiştir. Eğitim sürecinde tüm katmanlarda erken durdurma mekanizması devreye girmiş, ortalama eğitim döngüsü 9,2 ve toplam eğitim süresi 221,0 dakika olarak kaydedilmiştir. Farklı toplu işlem boyutları ve 5 katlı çapraz doğrulama sonuçları Ek 5'te sunulmuş olup, en yüksek test makro F1 değeri 0,8781, ağırlıklı F1 değeri ise 0,9260 olarak elde edilmiştir.

Modelin sınıf bazlı doğruluk dağılımı, Şekil 4.3'te verilen Normalize Edilmiş Karışıklık Matrisi ile incelenmiştir. Diyagonal üzerinde yer alan yüksek değerler, özellikle PERSON (0,92), NUMBER/DATE (0,97), SONG/ALBUM (0,93), MOVIE/TV (0,86) ve EVENT (0,85) etiketlerinde güçlü bir sınıflandırma başarısını ortaya koymaktadır. Buna karşılık LOCATION (0,74) ve ORGANIZATION (0,69) etiketlerinde doğruluk oranlarının görece daha düşük olduğu, bu sınıfların zaman zaman O veya birbirleriyle karıştığı görülmektedir. HEALTH etiketi ise 0,57 doğrulukla sınıflandırılmış ve çoğunlukla O sınıfına kaymıştır. Ayrıca RELATIONSHIP etiketi yalnızca 0,80 oranında doğru tahmin edilmiş, 0,20'si ise O etiketiyle karışmıştır. Bu durum, örnek sayısı düşük olan veya bağlamsal olarak benzer sınıflarda modelin daha fazla belirsizlik yaşadığını göstermektedir.



Şekil 4.3. Normalize Edilmiş Karışıklık Matrisi (mBERT, non-BIO)

Modelin kelime temsillerini iki boyutlu uzaya indirgenmiş hâli, Şekil 4.4'teki t-SNE Kümelendirme Görselleştirmesi ile sunulmuştur. Görsel incelendiğinde, PERSON, LOCATION, MOVIE/TV ve SONG/ALBUM gibi yaygın etiketlerin net kümeler hâlinde ayrıştığı, nadir görülen HEALTH, RELATIONSHIP ve SOCIAL_MEDIA gibi sınıfların ise daha dağınık bir dağılım gösterdiği gözlemlenmektedir. Bu durum, seyrek sınıflarda veri çeşitliliğinin artırılmasının modelin ayırım gücünü geliştirebileceğini öngörülmektedir.



Şekil 4.4. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (mBERT, non-BIO)

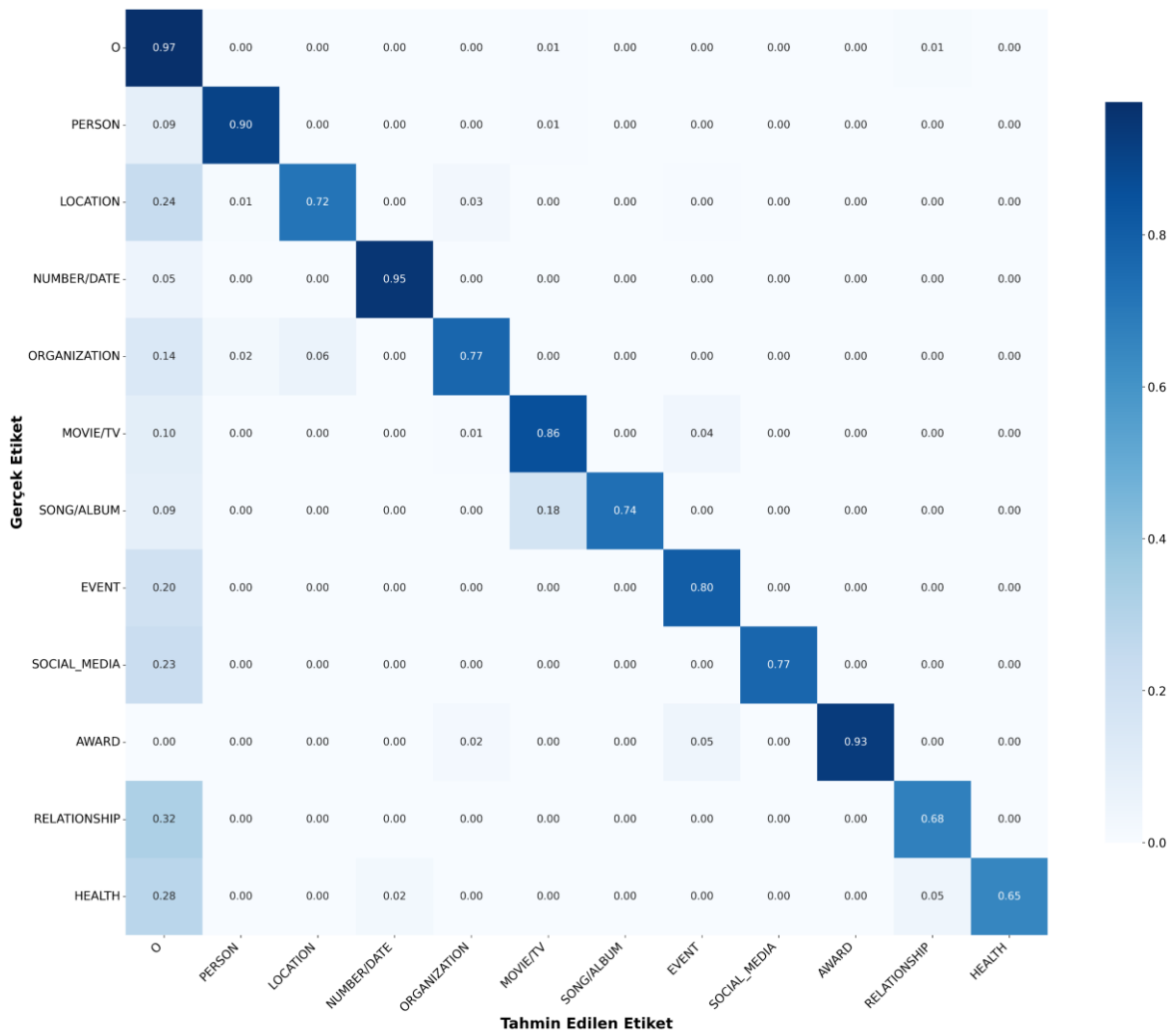
Sonuç olarak, mBERT modeli non-BIO formatta yaygın sınıflarda yüksek doğruluk, nadir sınıflarda ise temkinli bir tahmin stratejisi sergilemiştir. Karışıklık matrisi ve t-SNE dağılımı birlikte değerlendirildiğinde, modelin güvenilir tahminler ürettiği, ancak az görülen veya anlam bakımından birbirine yakın etiketler arasındaki ayrımın iyileştirilmesi gerektiği anlaşılmaktadır.

4.2.3. XLM-R (non-BIO)

Facebook AI tarafından geliştirilen çok dilli XLM-RoBERTa modeli, non-BIO etiketleme biçimiyle yapılan deneylerde yüksek doğruluk ve dengeli bir sınıf başarımı göstermiştir. Eğitim sürecinde tüm katmanlarda erken durdurma mekanizması etkin biçimde devreye girmiş; ortalama eğitim döngüsü sayısı 11,7, toplam eğitim süresi ise yaklaşık 257,3 dakika olarak kaydedilmiştir. Bu durum, modelin diğer modellere kıyasla daha uzun bir öğrenme süresine ihtiyaç duyduğunu ancak sonuçta kararlı ve güvenilir tahminler ürettiğini göstermektedir. Farklı toplu işlem boyutları ve çapraz doğrulama sonuçları Ek 6'da

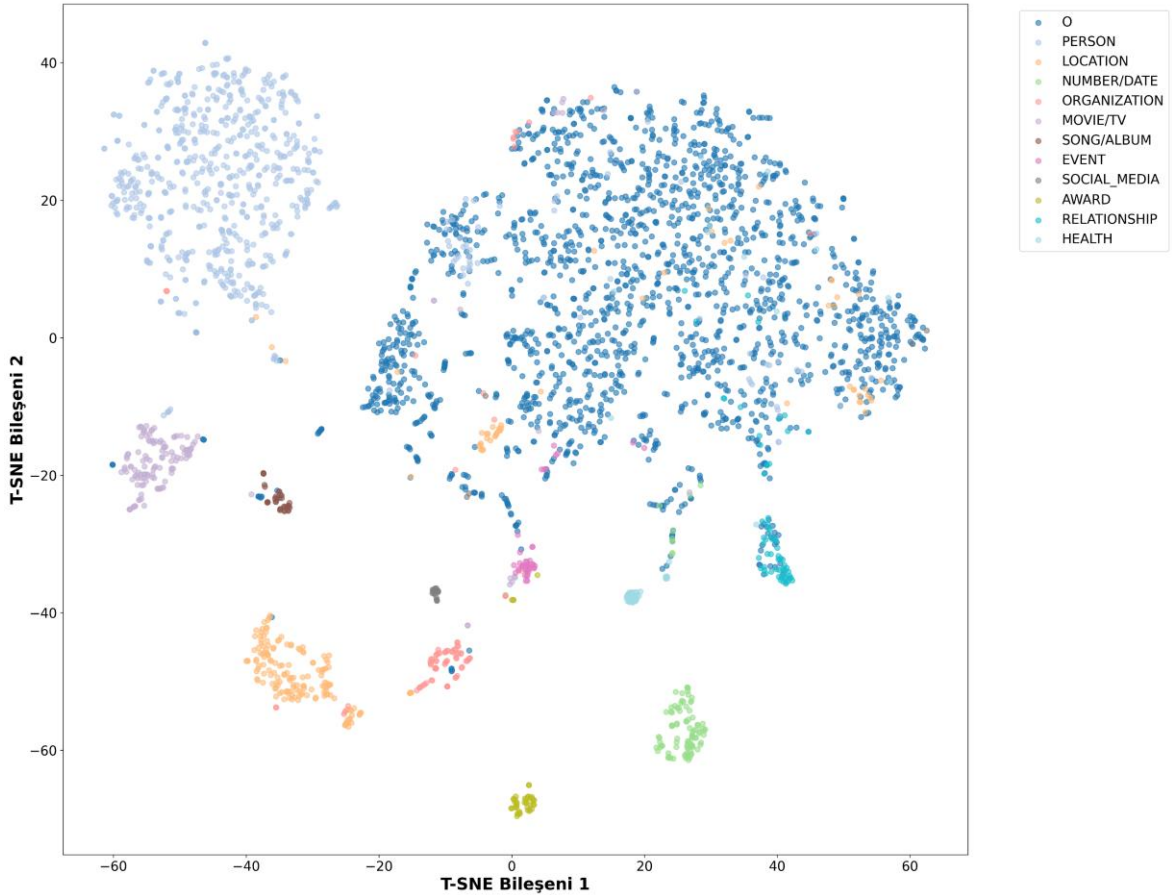
sunulmuştur. En yüksek test makro F1 değeri 0,8528 olup bu sonuç 8’lik toplu işlem boyutuna sahip üçüncü katmanda elde edilmiştir.

Modelin sınıf bazlı başarımı, Şekil 4.5’te verilen Normalize Edilmiş Karışıklık Matrisi üzerinden incelenmiştir. Matriste diyagonal üzerindeki yüksek değerler, modelin O (0,97), NUMBER/DATE (0,95), MOVIE/TV (0,86) ve PERSON (0,90) gibi yaygın sınıflarda yüksek doğruluk elde ettiğini göstermektedir. Bununla birlikte LOCATION (0,72) ve SONG/ALBUM (0,74) etiketlerinde doğruluk oranlarının daha düşük olduğu, bu sınıfların zaman zaman O veya MOVIE/TV gibi diğer etiketlerle karıştığı görülmektedir. RELATIONSHIP (0,68) ve HEALTH (0,65) etiketleri ise hem düşük örnek sayısına sahip hem de bağlamsal olarak diğer sınıflara yakın olduğundan, doğru sınıflandırma oranları görece düşük kalmıştır.



Şekil 4.5. Normaliz Edilmiş Karışıklık Matrisi (XLM-R, non-BIO)

Modelin ürettiği kelime temsillerinin iki boyutlu uzaydaki dağılımı, Şekil 4.6’da yer alan t-SNE Kümelenme Görselleştirme ile sunulmuştur. Görsel incelendiğinde, PERSON, MOVIE/TV ve NUMBER/DATE gibi yaygın sınıfların belirgin kümeler hâlinde ayrıştığı; buna karşın HEALTH, RELATIONSHIP ve SOCIAL_MEDIA gibi az görülen etiketlerin dağılımının daha dağınık olduğu görülmektedir. Bu durum, nadir etiketlerde veri çeşitliliğinin artırılmasının modelin ayırım gücünü iyileştirebileceği öngörülmektedir.



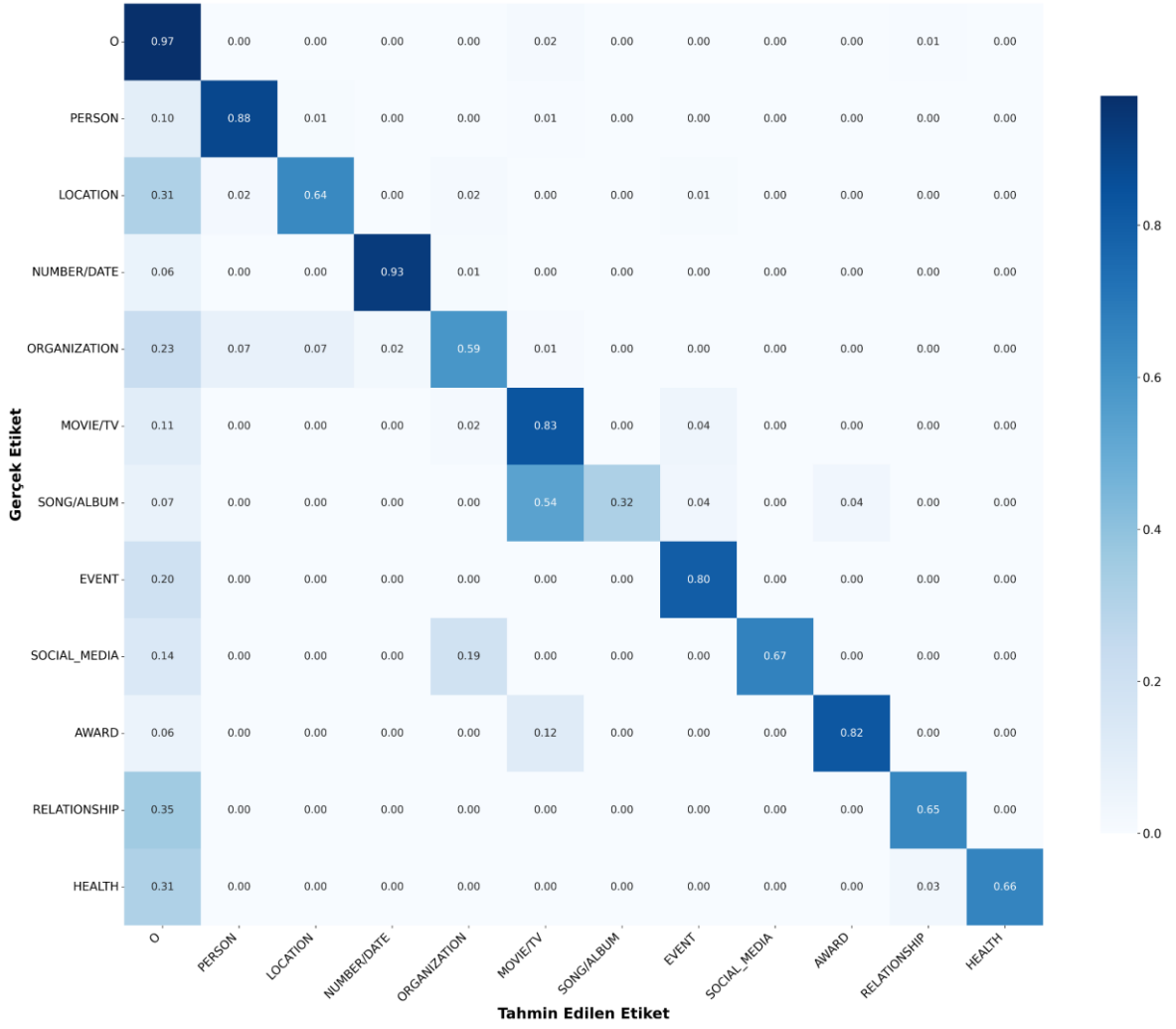
Şekil 4.6. Etiket Bazlı t-SNE Kümelenme Görselleştirme (XLM-R, non-BIO)

Sonuç olarak, XLM-R modeli non-BIO formatta yaygın etiketlerde yüksek doğrulukla çalışırken, az görülen veya anlam bakımından yakın etiketlerde zaman zaman karışıklıklar yaşamaktadır. Karışıklık matrisi ve t-SNE görselleştirme birlikte değerlendirildiğinde, modelin genel tahminlerinin güvenilir olduğu; ancak düşük örnek sayısına sahip sınıflarda ek veri ve iyileştirme ile başarımın artırılabilceği anlaşılmaktadır.

4.2.4. DistilBERTurk (non-BIO)

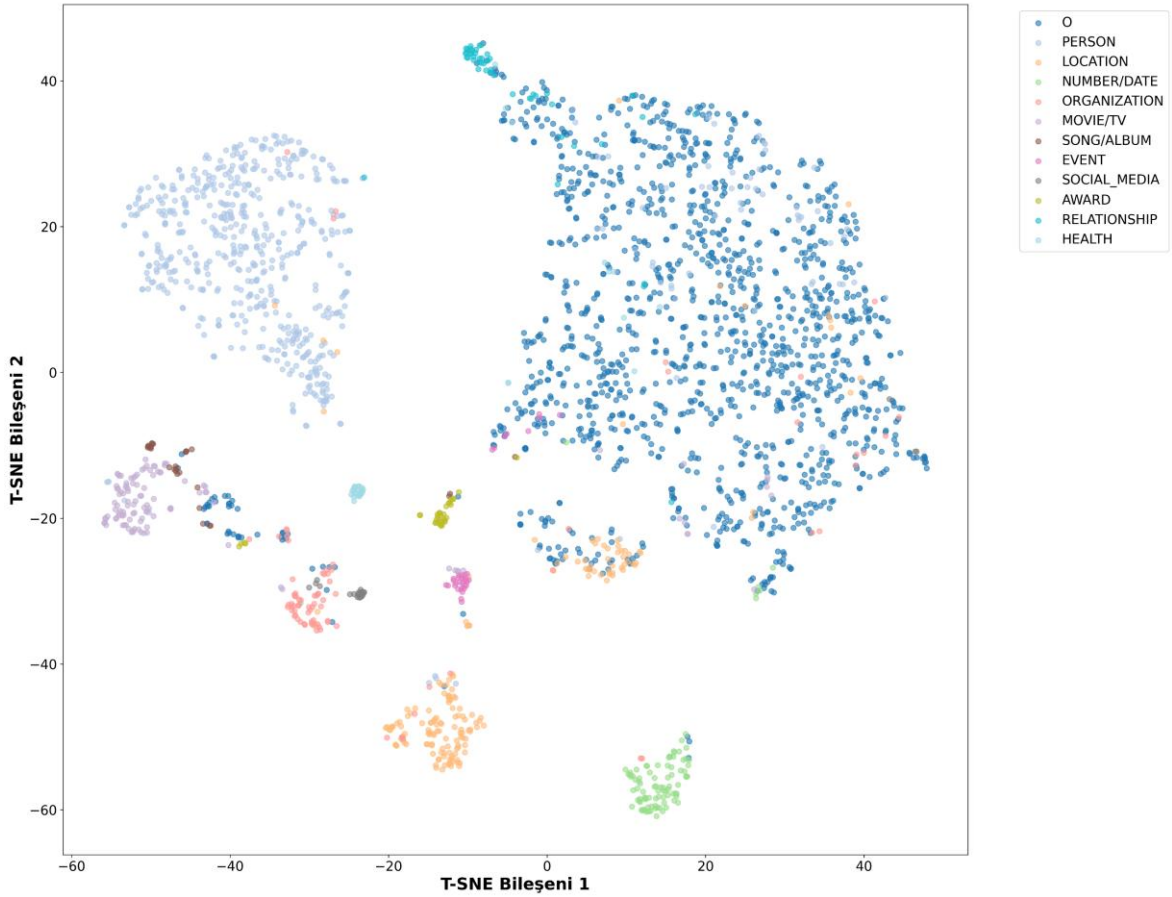
DistilBERTurk modeli, Türkçe için özel olarak eğitilmiş ve BERT mimarisinin daha kompakt bir versiyonudur. Bu çalışmada non-BIO etiketleme biçimiyle test edilmiş; tüm katmanlarda erken durdurma mekanizması etkin biçimde devreye girmiştir. Ortalama eğitim döngüsü sayısı 14,7, toplam eğitim süresi ise 125,7 dakika olarak ölçülmüştür. Model, görece kısa sürede eğitilebilmesine rağmen daha fazla eğitim döngüsü ile öğrenme sürecini tamamlamış ve istikrarlı bir performans sergilemiştir. Farklı toplu işlem boyutları ile elde edilen ayrıntılı sonuçlar Ek 7’de sunulmuştur. En yüksek test makro F1 değeri 0,8202 olup, bu sonuç 8’lik toplu işlem boyutunda dördüncü katmanda elde edilmiştir.

Modelin sınıf bazlı doğruluk düzeyleri Şekil 4.7’de yer alan Normalize Edilmiş Karışıklık Matrisi ile incelenmiştir. Görselde O (0,97), NUMBER/DATE (0,93), MOVIE/TV (0,83) ve EVENT (0,80) etiketlerinde yüksek doğruluk oranları dikkat çekmektedir. Ancak ORGANIZATION (0,59), LOCATION (0,64) ve özellikle SONG/ALBUM (0,54) etiketlerinde başarı düzeyi belirgin şekilde düşmektedir. SONG/ALBUM sınıfının 0,32 oranında MOVIE/TV ile karışması, bu iki sınıf arasında ayırt etme zorluğunu ortaya koymaktadır. RELATIONSHIP (0,65) ve HEALTH (0,66) gibi az sayıda örneğe sahip etiketlerde ise hataların önemli bir kısmının O sınıfına yönlendiği görülmektedir.



Şekil 4.7. Normalize Edilmiş Karışıklık Matrisi (DistilBERTTurk, non-BIO)

Modelin kelime temsillerinin iki boyutlu uzaydaki dağılımı, Şekil 4.8’de sunulan t-SNE Kümelendirme Görselleştirilmesi ile değerlendirilmiştir. Görsel incelendiğinde, NUMBER/DATE, MOVIE/TV ve EVENT gibi yaygın sınıfların kendi içinde yoğun kümeler oluşturduğu; SONG/ALBUM, RELATIONSHIP ve HEALTH gibi sınıfların ise daha dağınık bir yerleşime sahip olduğu görülmektedir. Bu dağınıklık, modelin bu etiketleri ayırt etmede daha fazla güçlük yaşadığını ve örnek çeşitliliğinin artırılmasıyla iyileştirme yapılabileceğini ortaya koymaktadır.



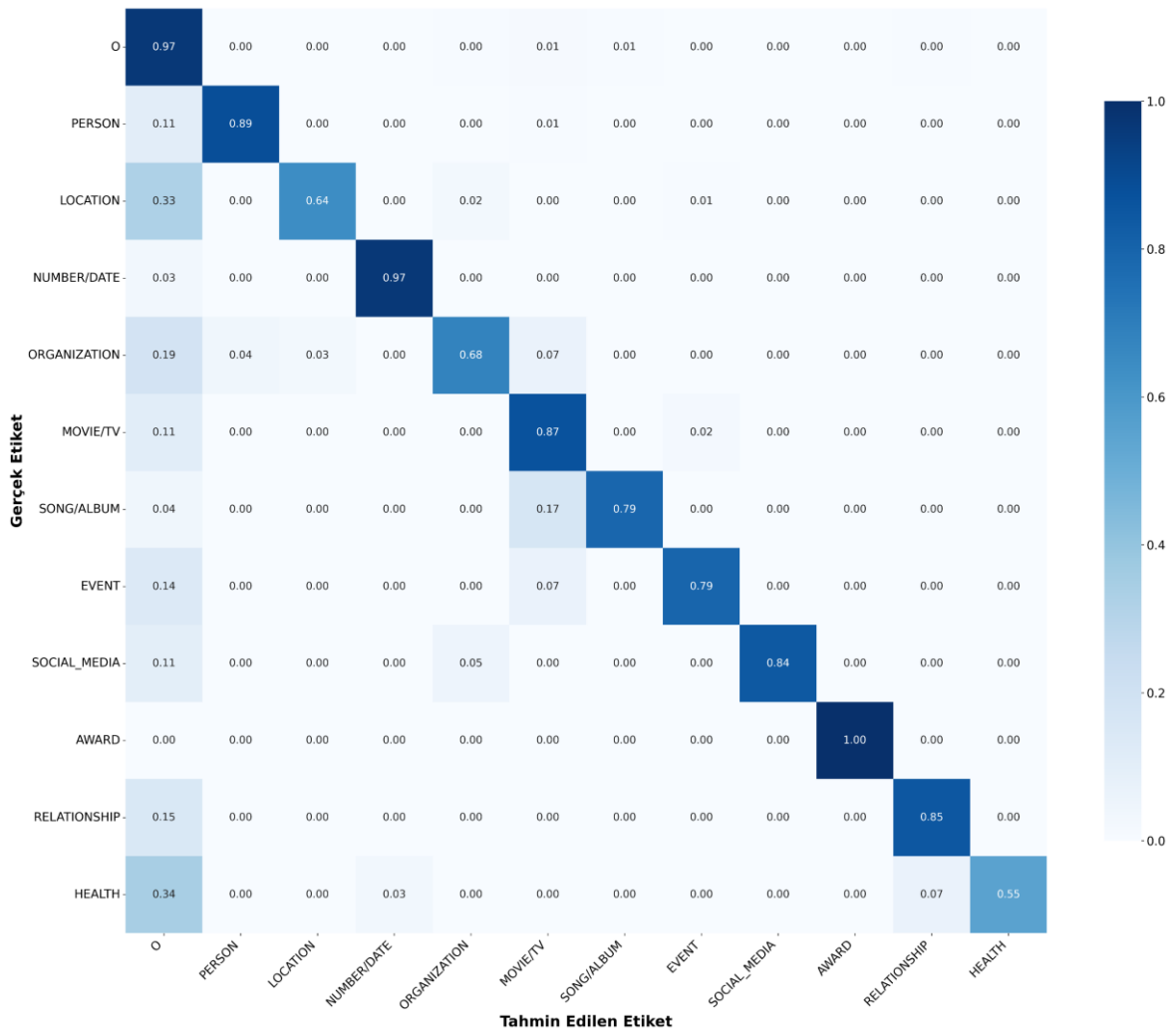
Şekil 4.8. Etiket Bazlı t-SNE Kümelenme Görselleştirme (DistilBERTTurk, non-BIO)

Sonuç olarak, DistilBERTTurk modeli yaygın etiketlerde yüksek doğruluk ve tutarlılık sunarken, az görülen veya içerik bakımından benzer sınıflarda ayırım gücü düşmektedir. Karışıklık matrisi ve t-SNE sonuçları birlikte değerlendirildiğinde, modelin özellikle SONG/ALBUM ve RELATIONSHIP gibi sınıflarda daha fazla veri ve ön işleme stratejileri ile desteklenmesinin başarıyı artırabileceği anlaşılmaktadır.

4.2.5. ConvBERTTurk (non-BIO)

ConvBERTTurk modeli, non-BIO etiketleme biçiminde yapılan deneylerde genel olarak dengeli ve güçlü bir performans sergilemiştir. Eğitim sürecinde tüm katmanlarda erken durdurma mekanizması etkin şekilde çalışmış; ortalama eğitim döngüsü sayısı 12,5 olarak ölçülmüştür. En kısa durdurma 9, en uzun durdurma ise 18 eğitim döngüsünde gerçekleşmiş ve tüm deneyler ortalama 223,1 dakikada tamamlanmıştır. Farklı toplu işlem boyutları (8, 16, 32) ve 5 katlı çapraz doğrulama ile elde edilen ayrıntılı sonuçlar Ek 8’de sunulmuştur. En yüksek test makro F1 değeri 0,8594 olup, bu sonuç 8’lik toplu işlem boyutunda birinci katmanda elde edilmiştir.

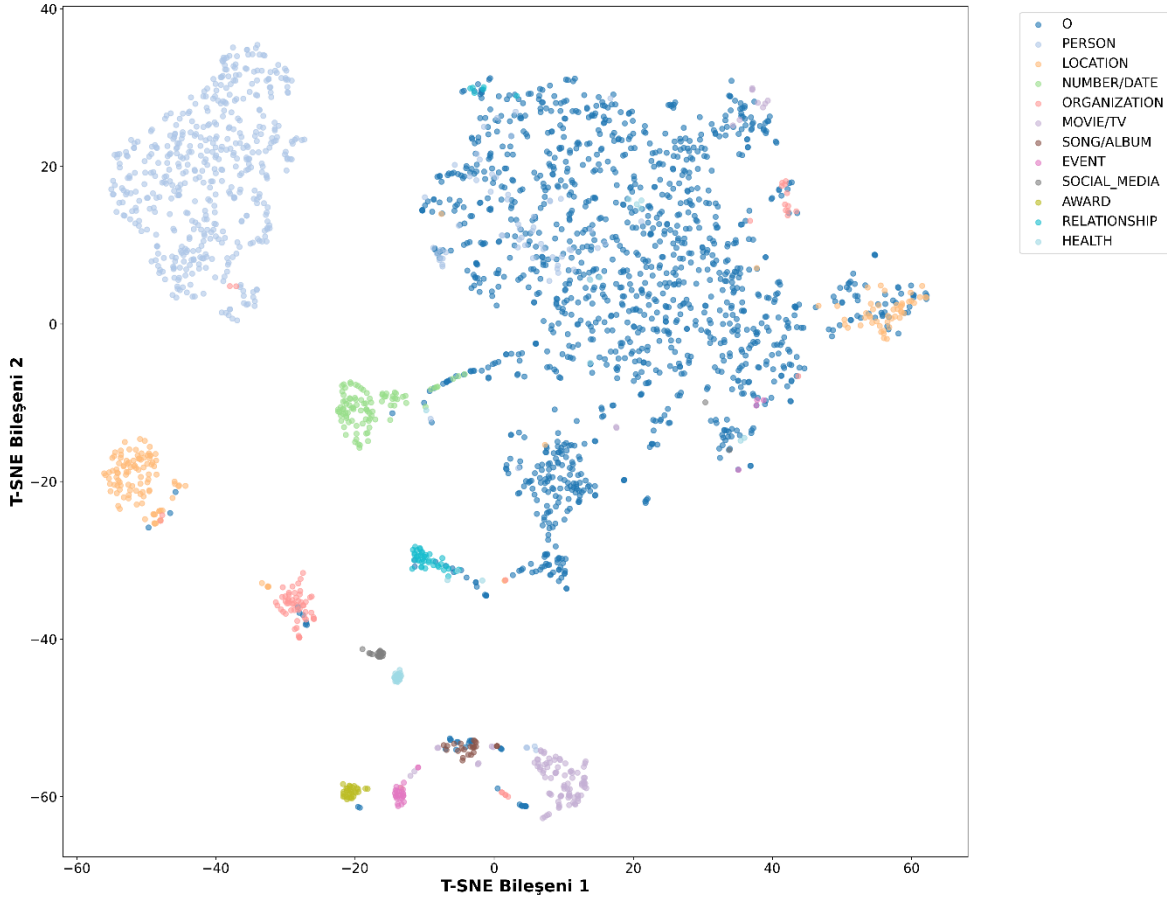
Modelin sınıf bazlı doğruluk oranları, Şekil 4.9’da yer alan Normalize Edilmiş Karışıklık Matrisi üzerinden değerlendirilmiştir. Görselde O (0,97), NUMBER/DATE (0,97), MOVIE/TV (0,87) ve SOCIAL_MEDIA (0,84) etiketlerinde yüksek doğruluk oranları dikkat çekmektedir. PERSON (0,89) ve AWARD (1,00) etiketlerinde de güçlü performans gözlemlenmiştir. Buna karşın HEALTH (0,55), LOCATION (0,64) ve ORGANIZATION (0,68) etiketlerinde doğruluk seviyeleri görece daha düşük olup, bu sınıfların zaman zaman O ya da içerik bakımından yakın diğer etiketlerle karıştığı görülmektedir. SONG/ALBUM (0,79) ve RELATIONSHIP (0,85) etiketlerinde ise doğru tahmin oranları görece yüksek olsa da, belirli bir oranda karışma yaşandığı anlaşılmaktadır.



Şekil 4.9. Normalize Edilmiş Karışıklık Matrisi (ConvBERTurk, non-BIO)

Modelin kelime temsillerinin iki boyutlu uzaydaki dağılımı, Şekil 4.10’da sunulan t-SNE Kümelendirme Görselleştirmesi ile incelenmiştir. Görselde PERSON, LOCATION, NUMBER/DATE ve MOVIE/TV etiketlerinin net kümeler hâlinde ayrıştığı; HEALTH,

RELATIONSHIP, EVENT ve SOCIAL_MEDIA gibi sınıfların ise daha dağınık bir şekilde konumlandığı görülmektedir. Özellikle HEALTH ve RELATIONSHIP etiketlerinin diğer sınıflara yakın dağılım göstermesi, modelin bu sınıflar arasındaki ayrımı yapmakta zorlandığını ortaya koymaktadır.



Şekil 4.10. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (ConvBERTurk, non-BIO)

Sonuç olarak, ConvBERTurk modeli non-BIO formatta yaygın sınıflarda yüksek doğruluk sunarken, az görülen veya birbirine yakın anlam taşıyan sınıflarda zaman zaman karışıklıklar yaşamaktadır. Karışıklık matrisi ve t-SNE görselleştirmesi birlikte değerlendirildiğinde, modelin genel tahminlerinin güvenilir olduğu; ancak düşük örnek sayısına sahip sınıflarda ek veri ve iyileştirme ile başarımın artırılabilceği anlaşılmaktadır.

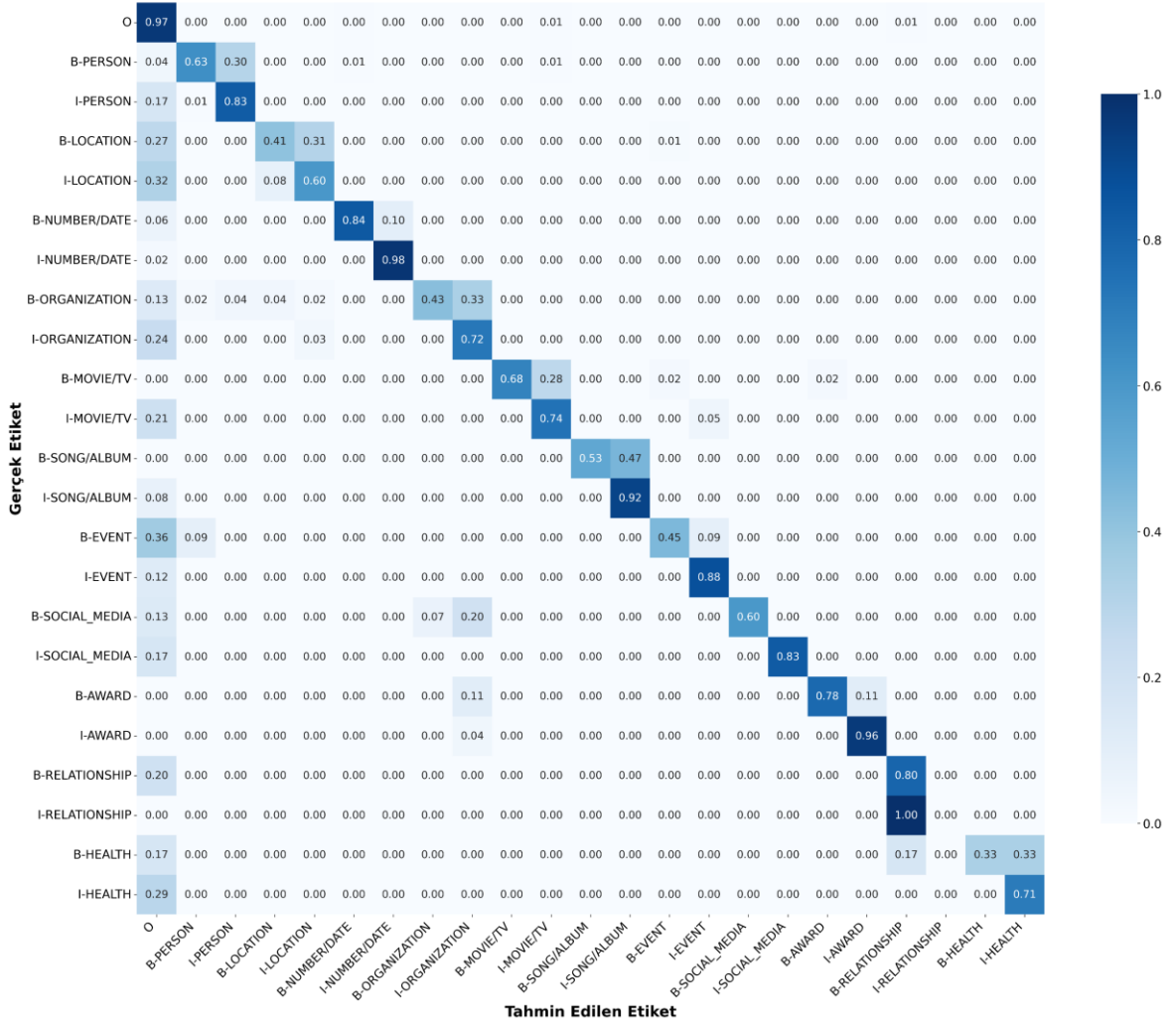
4.2.6. BERTurk (BIO)

BERTurk modeli, BIO etiketleme biçimiyle yapılan deneylerde genel olarak istikrarlı bir performans sergilemiştir. Eğitim sürecinde tüm alt parça eğitimlerinde erken durdurma mekanizması etkin biçimde çalışmış, ortalama gerçek döngü sayısı 10,7 olarak ölçülmüş ve

toplam eğitim süresi 183,6 dakika olarak hesaplanmıştır. Bu veriler, modelin öğrenme sürecinde dengeli ilerlediğini ve aşırı öğrenme eğilimine girmeden süreci tamamladığını göstermektedir.

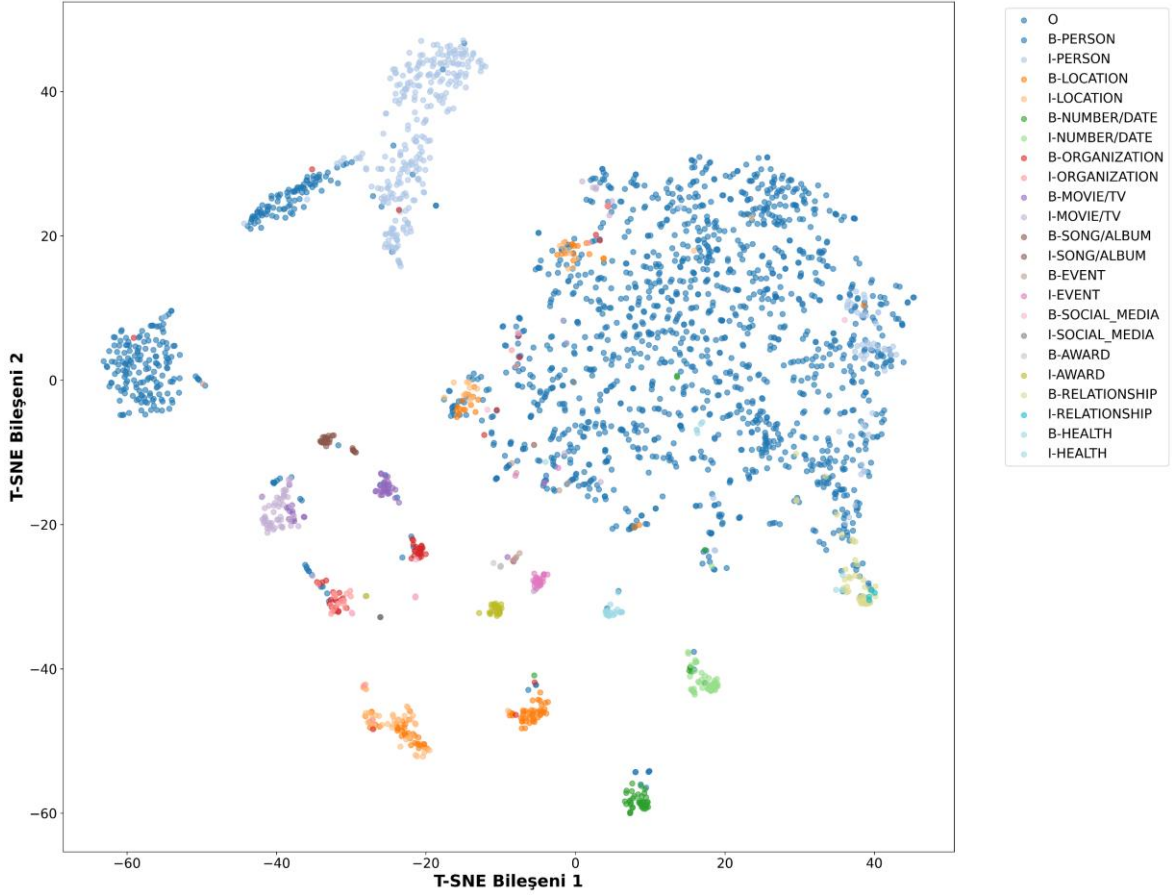
Ek 9’da sunulan ayrıntılı sonuçlara göre, en yüksek test makro F1 değeri 0,7194 ile 16 birimlik toplu işlem boyutunda, 5. alt parçada elde edilmiştir. Diğer deneylerde makro F1 değerleri 0,6893 ile 0,7194 arasında değişirken, ağırlıklı F1 değerleri 0,8297 ile 0,8425 aralığında kalmıştır. Bu durum, modelin özellikle yaygın görülen etiketlerde daha yüksek başarı sağladığını ortaya koymaktadır.

Modelin sınıf bazlı doğruluk dağılımı Şekil 4.11’de sunulan normalize edilmiş karışıklık matrisi ile incelenmiştir. Grafik, modelin O, B-NUMBER/DATE ve I-NUMBER/DATE gibi yaygın etiketlerde yüksek doğruluk oranlarına ulaştığını; buna karşılık B-ORGANIZATION ile I-ORGANIZATION veya B-LOCATION ile I-LOCATION gibi yapısal olarak benzer etiketlerde belirgin çapraz tahminler yaptığını göstermektedir. Ayrıca B-HEALTH ve I-HEALTH etiketlerinde, tahminlerin önemli bir kısmının O etiketine yöneldiği görülmektedir.



Şekil 4.11. Normalize Edilmiş Karışıklık Matrisi (BERTurk, BIO)

Modelin ürettiği kelime temsillerinin iki boyutlu uzaydaki dağılımı Şekil 4.12'deki t-SNE görselleştirmesi ile sunulmuştur. Görsel incelendiğinde, B-PERSON ve I-PERSON etiketlerinin yoğun ve ayrık kümeler oluşturduğu, B-NUMBER/DATE ve I-NUMBER/DATE etiketlerinin de net biçimde ayrıldığı görülmektedir. Ancak B-SONG/ALBUM ile I-SONG/ALBUM veya B-HEALTH ile I-HEALTH gibi bazı çiftlerin kümelerinin kısmen iç içe geçtiği dikkat çekmektedir. Bu durum, modelin bazı sınıflarda iç ve dış etiket ayırımında zorluk yaşadığını göstermektedir.



Şekil 4.12. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (BERTurk, BIO)

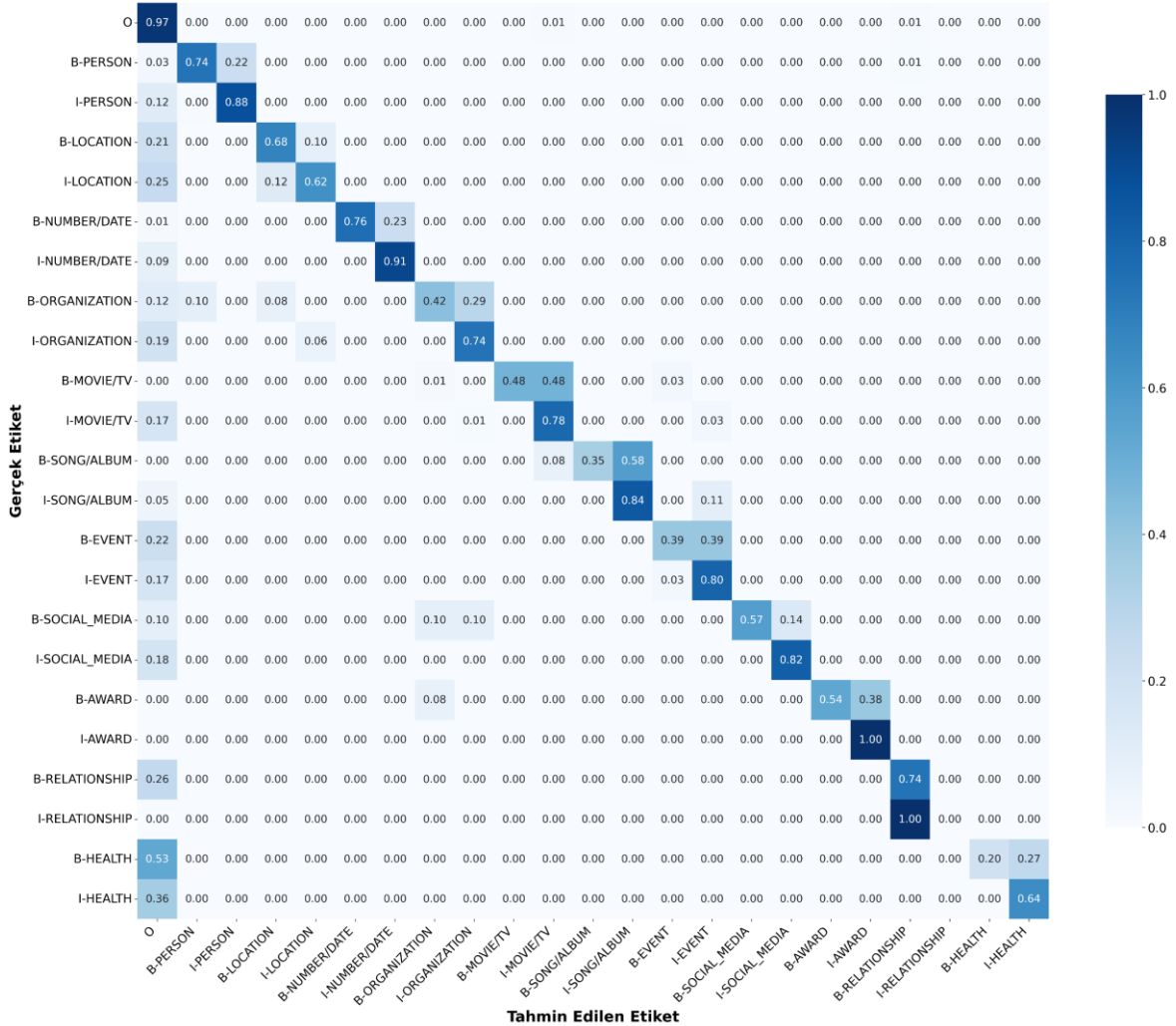
Sonuç olarak, BERTurk modeli BIO etiketleme yapısında güçlü ayrımlar yapabildiği etiketlerin yanı sıra, özellikle B-I etiket çiftlerinde ve düşük örnek sayısına sahip sınıflarda hata eğilimi göstermektedir. Karışıklık matrisi ve t-SNE görselleştirmesi birlikte değerlendirildiğinde, modelin başarılı olduğu alanlar ve geliştirilmesi gereken yönler açık biçimde ortaya çıkmaktadır.

4.2.7. mBERT (BIO)

Google tarafından geliştirilen çok dilli BERT modeli, BIO etiketleme formatı kullanılarak değerlendirilmiştir. Eğitim sürecinde üç farklı toplu işlem boyutunda (8, 16, 32) ve 15 alt parça üzerinde gerçekleştirilen tüm deneylerde erken durdurma mekanizması etkili biçimde çalışmıştır. Ortalama gerçek yineleme sayısı 11,5 olup, en kısa eğitim 7 yinelemede, en uzun eğitim ise 16 yinelemede tamamlanmıştır. Toplam eğitim süresi yaklaşık 281,3 dakika olarak ölçülmüştür. Ayrıntılı sonuçlar EK 10'da tablo hâlinde sunulmaktadır.

Modelin son test kümesinde ürettiği tahminlerin doğruluk dağılımı, Şekil 4.13'te verilen normalize edilmiş karışıklık matrisi ile görselleştirilmiştir. Bu matris, mBERT'in

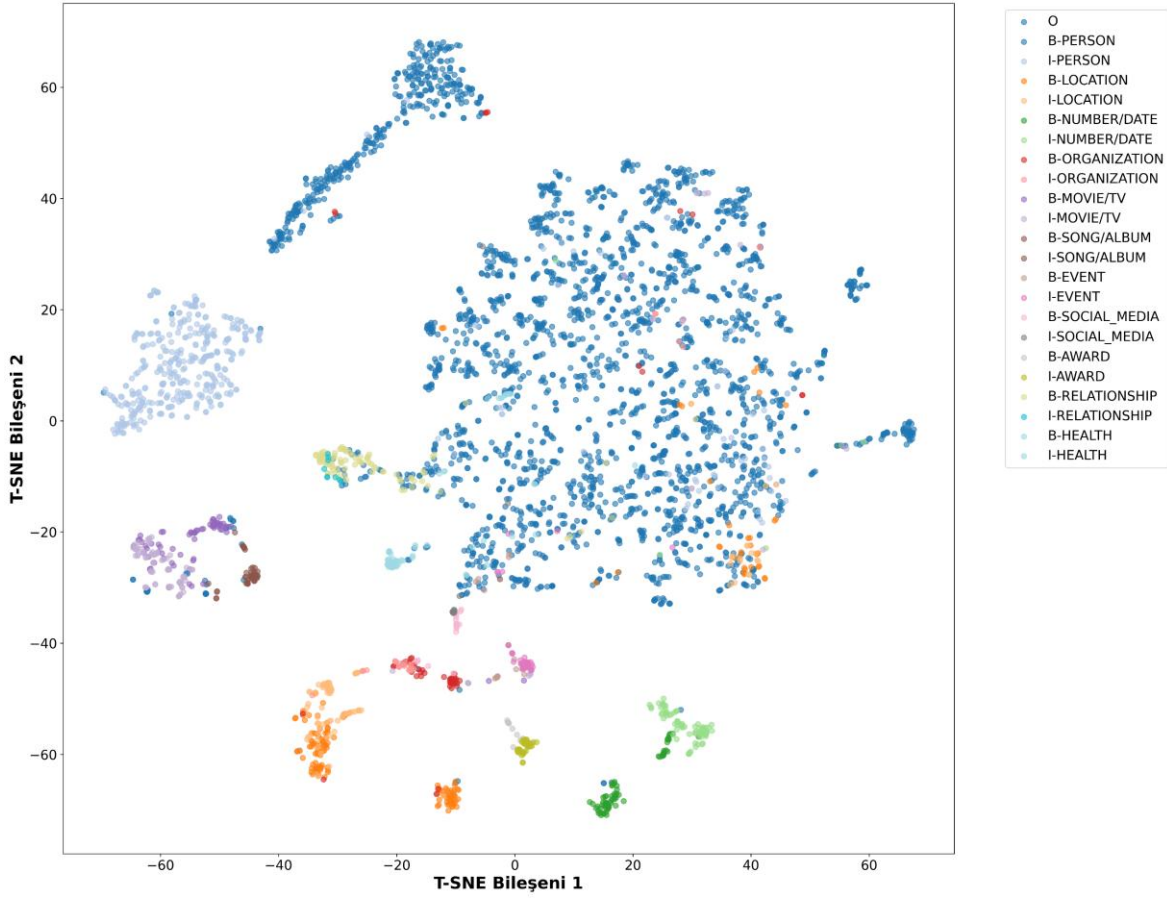
özellikle B-PERSON (0,74) ve I-PERSON (0,88) etiketlerinde yüksek doğruluk sağladığını ortaya koymaktadır. Ancak bazı etiketlerde belirgin karışmalar gözlemlenmiştir. Örneğin, I-HEALTH etiketli örneklerin 0,53'ünün B-PERSON, 0,36'sının ise B-HEALTH olarak tahmin edilmesi, sağlıkla ilgili varlıkların biyografik içeriklerle karıştırıldığını göstermektedir. Benzer şekilde, B-SONG/ALBUM (0,58) ve I-SONG/ALBUM (0,84) etiketlerinde çapraz hataların yoğunlaştığı görülmektedir.



Şekil 4.13. Normaliz Edilmiş Karışıklık Matrisi (mBERT, BIO)

Şekil 4.14'te sunulan t-SNE kümeleme görselleştirmesi, modelin etiketler arası ayırım gücünü iki boyutlu uzayda ortaya koymaktadır. PERSON ve NUMBER/DATE gibi sık görülen etiketler belirgin kümeler oluştururken, HEALTH, RELATIONSHIP ve SOCIAL_MEDIA gibi az sayıda örneğe sahip etiketler daha dağınık bir dağılım sergilemiştir. Bu durum, nadir etiketlerin öğrenilmesinde modelin zorlandığını göstermektedir. Ayrıca, B-I etiket çiftleri (ör. B-RELATIONSHIP – I-RELATIONSHIP, B-

LOCATION – I-LOCATION) arasında yakın konumlanmalar, iç ve başlangıç etiketlerinin ayırt edilmesinde yaşanan güçlükleri yansıtmaktadır.



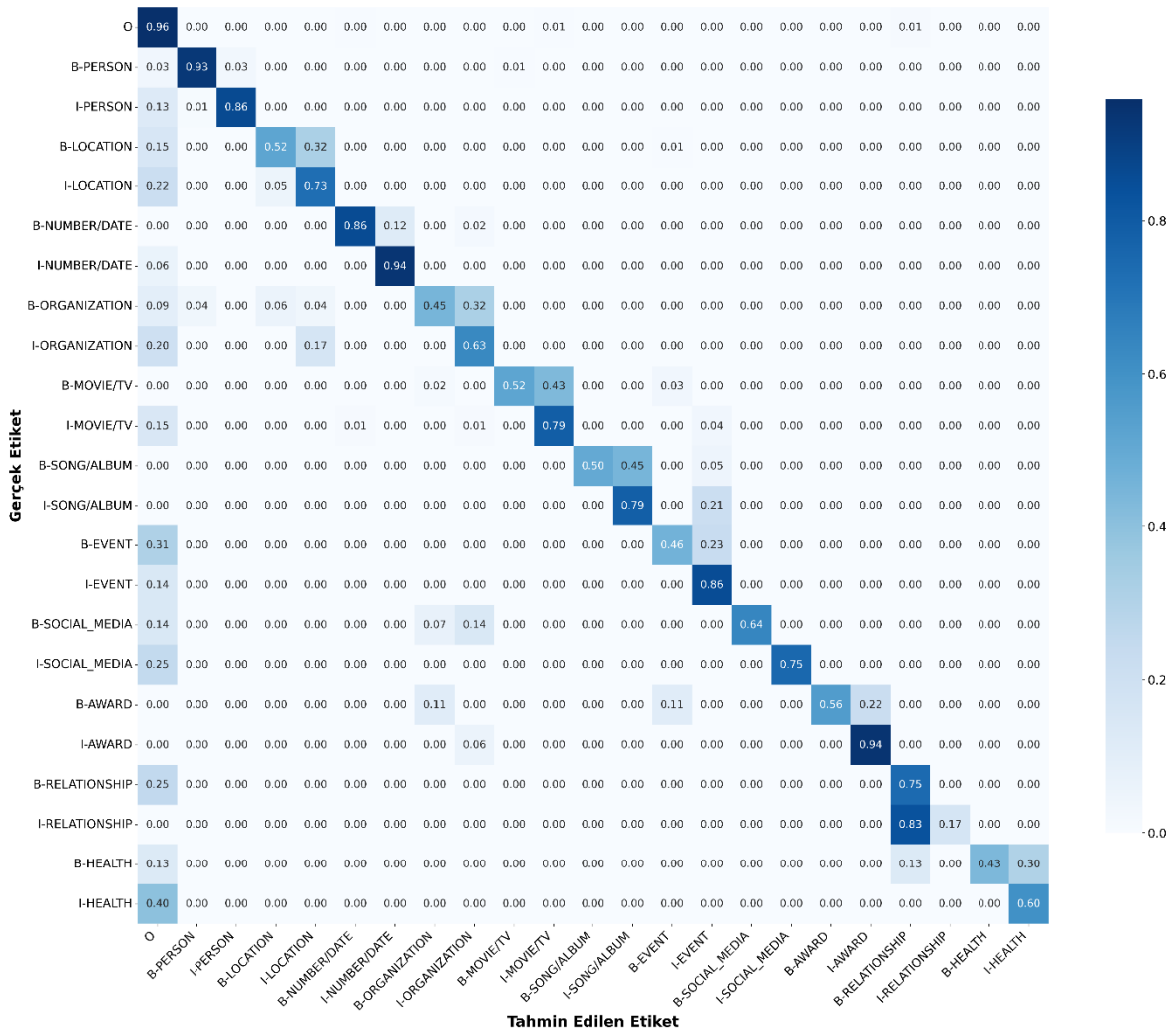
Şekil 4.14. Etiket Bazlı t-SNE Kümelenme Görselleştirme (mBERT, BIO)

Sonuç olarak, mBERT modeli BIO formatında belirli varlık türlerinde başarılı olsa da özellikle düşük örnek sayısına sahip etiketlerde ve B-I ayrımında hataya yatkın bir yapı sergilemiştir. Bu tablo, modelin çok dilli mimarisine rağmen Türkçenin biçimbilimsel özelliklerini tam anlamıyla yansıtmakta sınırlı kaldığını düşündürmektedir.

4.2.8. XLM-R (BIO)

XLM-R modeli, BIO etiketleme formatı ile 15 alt parça üzerinde değerlendirilmiştir. Tüm deneylerde erken durdurma mekanizması devreye girmiş, ortalama gerçek yineleme sayısı 13,3 olarak ölçülmüştür. En düşük yineleme sayısı 8, en yüksek ise 24'tür. Toplam eğitim süresi 379,9 dakika olarak gerçekleşmiştir. Ayrıntılı sonuçlar EK 11'de tablo halinde sunulmaktadır.

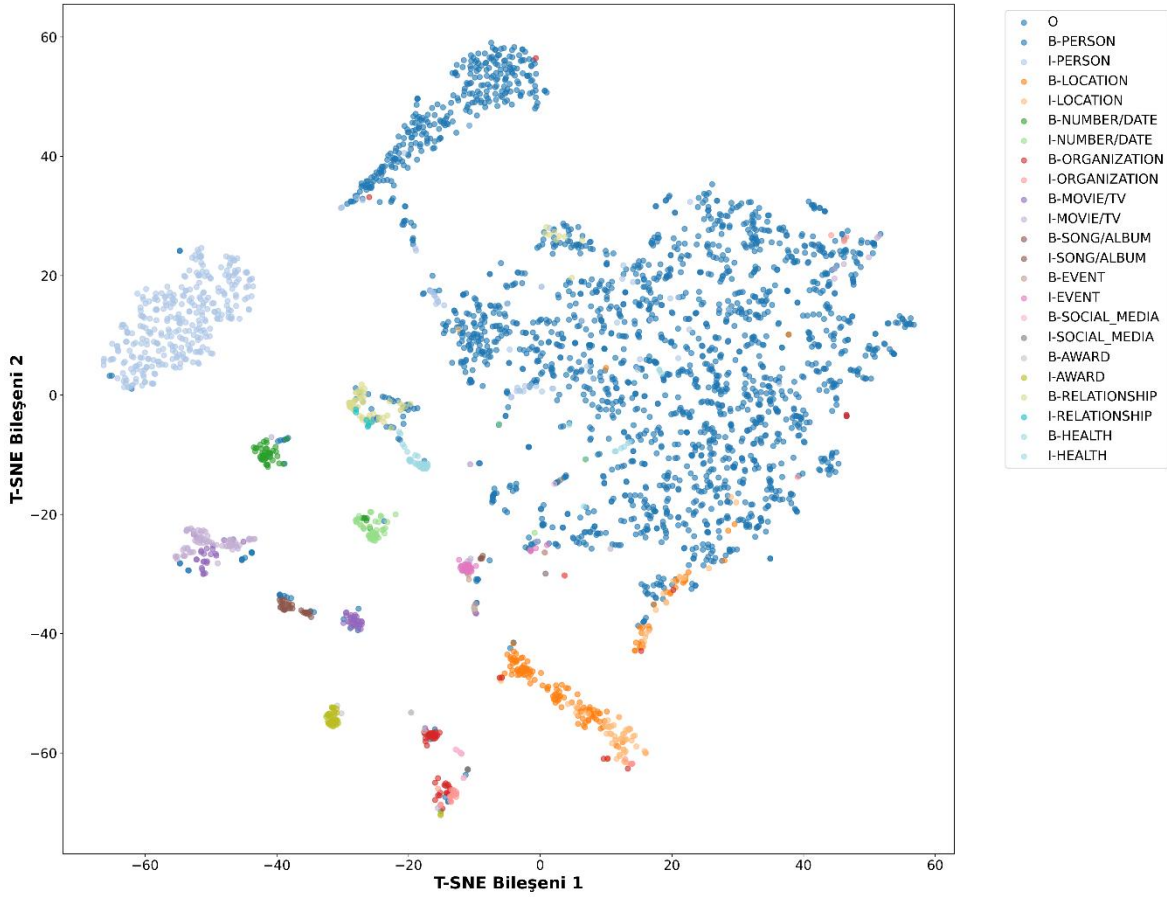
Etiket bazlı performans incelendiğinde, B-NUMBER/DATE, B-LOCATION ve B-PERSON gibi sık rastlanan sınıflarda başarılı sonuçlar elde edilmiştir. Buna karşın I-HEALTH, B-RELATIONSHIP ve B-SONG/ALBUM etiketlerinde belirgin performans düşüklüğü gözlenmiştir. Şekil 4.15’te sunulan normalize edilmiş karışıklık matrisi, hataların çoğunlukla benzer içerikli sınıflar arasında yoğunlaştığını ortaya koymaktadır. Örneğin I-HEALTH etiketli örneklerin 0,40’ı B-PERSON olarak tahmin edilmiştir. Benzer şekilde B-RELATIONSHIP etiketi, B-PERSON ve I-RELATIONSHIP ile yüksek oranda karıştırılmıştır. Bu durum, modelin özellikle anlam bakımından yakın etiketleri ayırt etmede zorlandığını göstermektedir.



Şekil 4.15. Normalizasyon Edilmiş Karışıklık Matrisi (XLM-R, BIO)

Modelin etiket ayırımını görsel olarak incelemek amacıyla yapılan t-SNE yerleştirmesi Şekil 4.16’da verilmiştir. Görselde, çoğu etiketin belirgin kümeler oluşturduğu görülmektedir. Ancak I-ORGANIZATION ve I-LOCATION gibi bazı sınıflarda

kümelenmenin daha dağınık olduğu dikkati çekmektedir. Bu durum, özellikle benzer bağlamlarda geçen varlıkların sınırlarının net olarak belirlenemediğini göstermektedir.



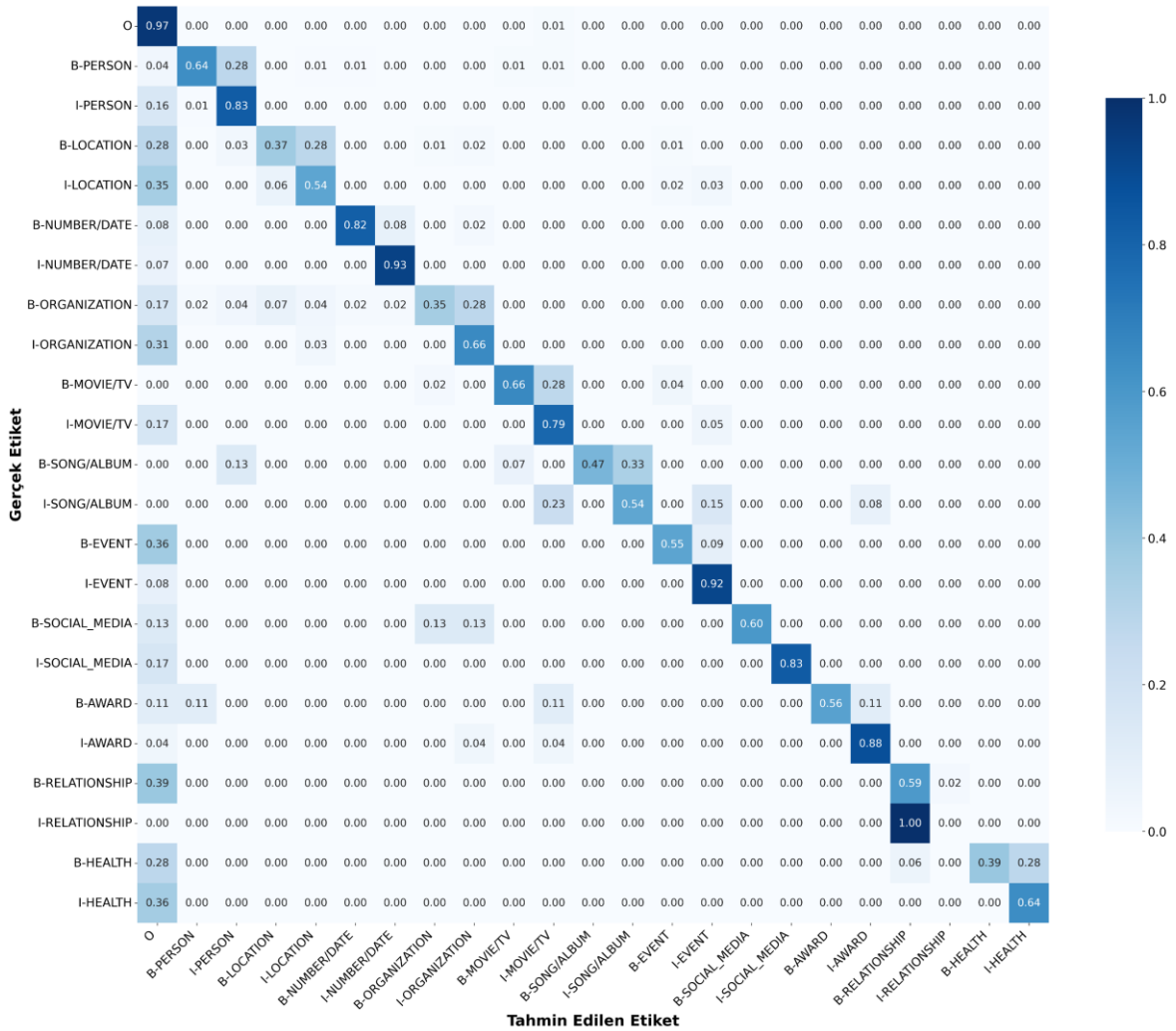
Şekil 4.16. Etiket Bazlı t-SNE Kümelenme Görselleştirme (XLM-R, BIO)

Genel olarak XLM-R modeli, BIO formatta dengeli bir performans sergilemiştir. Yaygın sınıflarda yüksek doğruluk sağlarken, daha az örneğe sahip etiketlerde ayırım gücü sınırlı kalmıştır. Veri çeşitliliğinin artırılması ve benzer sınıflar arası ayırım yeteneğinin güçlendirilmesi, modelin gelecekteki başarısını yükseltebilir.

4.2.9. DistilBERTurk (BIO)

DistilBERTurk modeli, BIO etiketleme formatıyla yapılan eğitimlerde erken durdurma mekanizmasını etkin şekilde kullanmıştır. On beş farklı deney sonucunda ortalama gerçek yinleme sayısı 15,1 olarak ölçülmüş, en kısa eğitim 9, en uzun eğitim ise 22 yinelemede tamamlanmıştır. Toplam eğitim süresi yaklaşık 466,0 dakika olarak kaydedilmiştir. Ayrıntılı eğitim sonuçları EK 12’de tablo hâlinde sunulmaktadır.

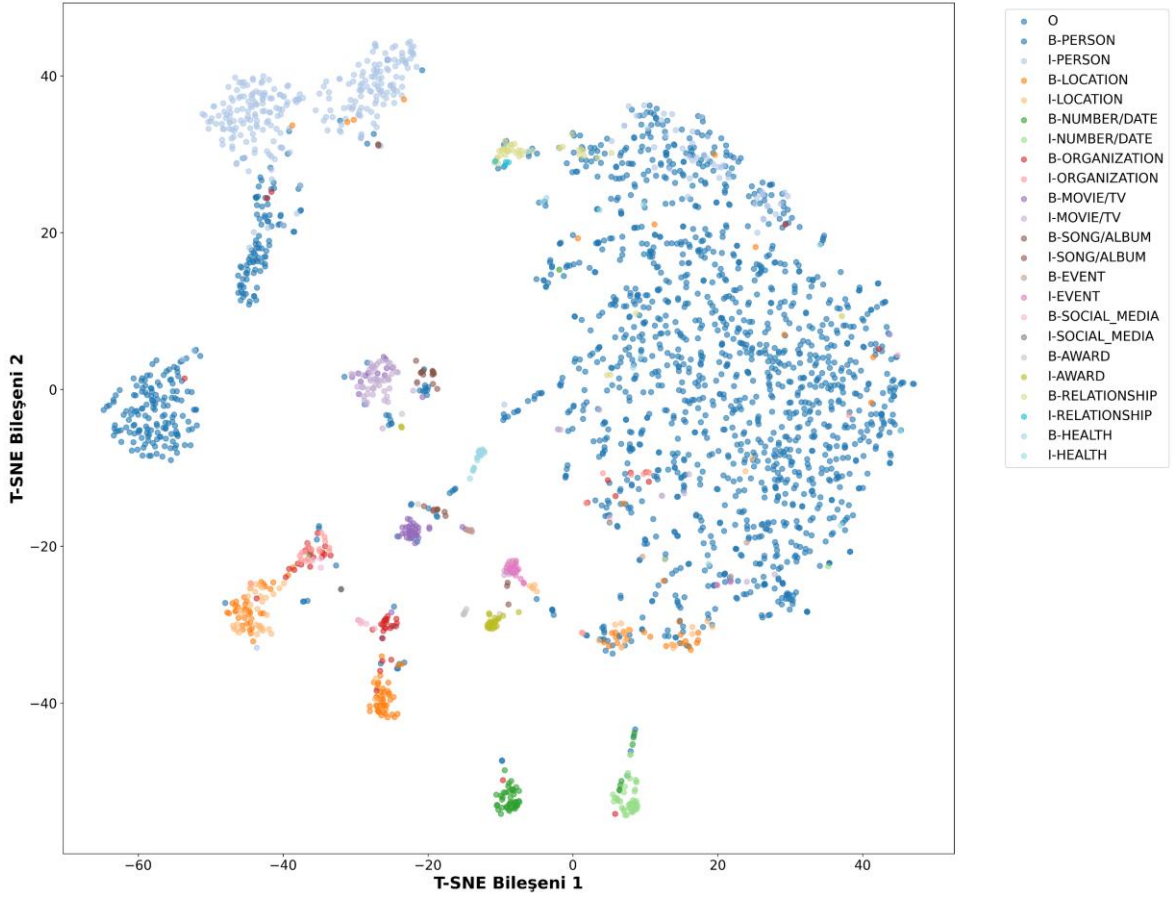
Modelin sınıf bazlı tahmin dağılımlarını analiz edebilmek amacıyla oluşturulan normalize edilmiş karışıklık matrisi Şekil 4.17’de yer almaktadır. Matris incelendiğinde, özellikle B-PERSON ve I-PERSON etiketlerinde karşılıklı karışıklıkların belirgin olduğu, B-RELATIONSHIP etiketinin ise çoğunlukla I-RELATIONSHIP olarak tahmin edildiği görülmektedir. I-HEALTH etiketinin yüksek oranda B-HEALTH sınıfına yönelmesi, BIO yapısındaki başlangıç ve iç ayrımının bazı durumlarda net öğrenilemediğini göstermektedir. Buna karşılık, B-SONG/ALBUM ve I-SONG/ALBUM gibi sınıflarda tahminlerin daha isabetli olduğu dikkati çekmektedir.



Şekil 4.17. Normalize Edilmiş Karışıklık Matrisi (DistilBERTTurk, BIO)

Modelin düşük boyutlu temsil uzayındaki karar sınırlarını anlamak için yapılan t-SNE kümelenme analizi Şekil 4.18’de sunulmuştur. Görsel incelendiğinde, B-PERSON, I-PERSON, B-NUMBER/DATE ve I-NUMBER/DATE etiketlerinin belirgin kümeler hâlinde ayrıştığı, ancak bazı sınıflarda kümelerin kısmen iç içe geçtiği gözlemlenmiştir.

Özellikle B-I ayrımlarının yapıldığı etiketlerde bu durumun daha belirgin olması, modelin bu tür etiket çiftlerinde ayırım yapma becerisinin sınırlı olabileceğini düşündürmektedir.



Şekil 4.18. Etiket Bazlı t-SNE Kümelenme Görselleştirmesi (DistilBERTurk, BIO)

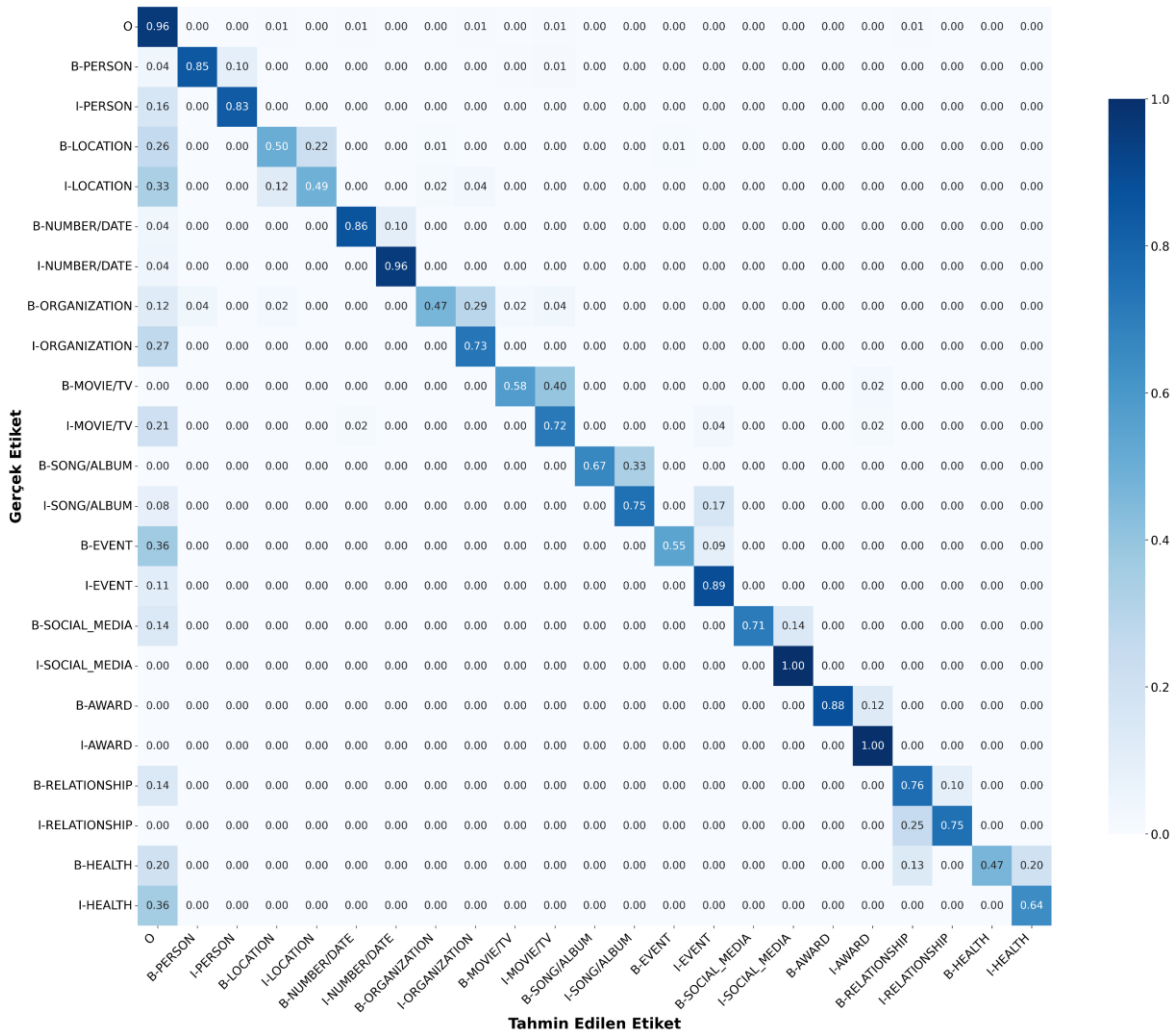
Genel olarak, DistilBERTurk modeli parametre açısından daha sade olmasına rağmen bazı varlık türlerinde yüksek doğruluk sağlamış, ancak BIO formatındaki etiket ayrımlarında belirli sınıflarda zorluk yaşamıştır. Bu nedenle, düşük kaynaklı dil senaryolarında ve karmaşık etiketleme görevlerinde dikkatli şekilde değerlendirilmesi gereken bir model olarak öne çıkmaktadır.

4.2.10. ConvBERTurk (BIO)

ConvBERTurk modeli, BIO etiketleme formatı altında değerlendirilmiştir. Eğitim süreci, üç farklı toplu işlem boyutunda (8, 16, 32) yürütülen toplam 15 alt deneyden oluşmuş ve tüm deneylerde erken durdurma mekanizması devreye girmiştir. Ortalama gerçek yineleme sayısı 15,1 olarak ölçülmüş; en düşük değer 9, en yüksek değer ise 25 olmuştur.

Toplam eğitim süresi yaklaşık 365,7 dakika olarak kaydedilmiştir. Her bir konfigürasyona ilişkin ayrıntılı sonuçlar EK 13’te tablo hâlinde sunulmaktadır.

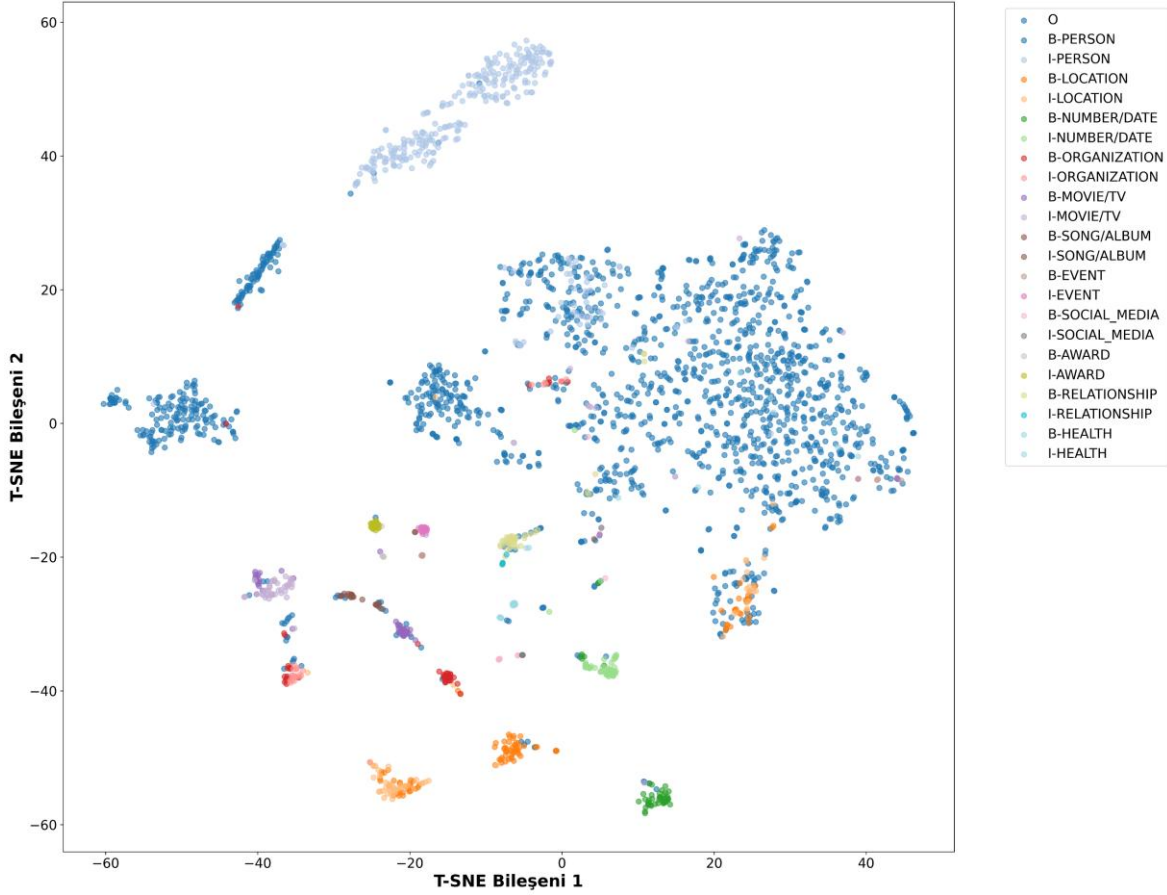
Modelin sınıf bazlı tahmin dağılımlarını analiz edebilmek amacıyla oluşturulan normalize edilmiş karışıklık matrisi Şekil 4.19’da verilmiştir. Matris incelendiğinde, I-RELATIONSHIP ve B-RELATIONSHIP etiketleri arasında yüksek oranda karışıklık olduğu görülmektedir. Benzer şekilde, I-HEALTH etiketinin B-HEALTH ile ve I-ORGANIZATION etiketinin B-ORGANIZATION ile sıklıkla karıştırıldığı tespit edilmiştir. Bu bulgular, modelin BIO formatında iç ve başlangıç etiket ayrımını her zaman net bir şekilde yapamadığını göstermektedir.



Şekil 4.19. Normalize Edilmiş Karışıklık Matrisi (ConvBERTurk, BIO)

Modelin karar sınırlarının görsel olarak incelenebilmesi için elde edilen t-SNE dağılımı Şekil 4.20’de sunulmuştur. Görselde, B-PERSON, B-LOCATION ve B-NUMBER/DATE gibi sık görülen sınıfların belirgin kümeler oluşturduğu gözlenmektedir.

Buna karşın, I-RELATIONSHIP ve I-HEALTH gibi bazı sınıfların kümelerinin birbirine oldukça yakın konumlandığı dikkat çekmektedir. Bu durum, belirli alt sınıflar arasında ayrım yapmanın model için daha güç olduğunu ortaya koymaktadır.



Şekil 4.20. Varlık Türlerine Göre t-SNE Dağılımı (ConvBERTurk, BIO)

Genel olarak ConvBERTurk, BIO etiketleme formatında tatmin edici bir performans sergilemekle birlikte, özellikle B-I etiket çiftlerinde yaşanan karışıklıklar, etiket ayrım gücünün geliştirilmesine yönelik ek çalışmalara ihtiyaç duyulduğunu göstermektedir.

4.3. Modeller Arası Karşılaştırmalı Performans Analizi

Bu çalışmada kullanılan beş farklı modelin hem BIO hem de non-BIO etiketleme formatları altında gösterdiği genel ve sınıf bazlı performanslar karşılaştırmalı olarak incelenmiştir. Her bir model, sabit hiper parametreler ve 5 katlı çapraz doğrulama yöntemi ile eğitilmiştir. Test verisi üzerinde doğruluk, makro kesinlik, makro duyarlılık, makro F1 ve ağırlıklı F1 ölçütleriyle değerlendirilmiştir. Sınıf bazlı detaylı sonuçlar EK 14 (non-BIO) ve EK 15 (BIO) eklerinde sunulmaktadır.

Tablo 4.1 incelendiğinde, non-BIO etiketleme formatında genel olarak daha yüksek doğruluk ve F1 skorlarının elde edildiği göze çarpmaktadır. Özellikle mBERT modeli, non-BIO formatta hem 0,93 doğruluk hem de 0,93 ağırlıklı F1 skoru ile öne çıkmaktadır. Ayrıca bu model, 0,87 makro F1 skoru ile BERTurk ile en yüksek ortalama değeri yakalamıştır. ConvBERTurk modeli de bu formatta oldukça dengeli sonuçlar vermiş, 0,91 doğruluk ve 0,86 makro F1 ile etkili bir performans sergilemiştir. Bunun yanında, XLM-R ve BERTurk modelleri de 0,85–0,87 aralığında makro F1 skorları üretmiştir. DistilBERTurk ise diğer modellere kıyasla bir miktar daha düşük kalmış ve non-BIO şemada 0,82 makro F1 ile sınırlı bir başarı sağlamıştır. Genel değerlendirme olarak non-BIO formatta modellerin genellikle daha istikrarlı sonuçlar verdiği hem doğruluk hem de F1 metriklerinin yüksekliğiyle desteklenmiştir. BIO etiketleme formatına geçildiğinde ise genel performans seviyelerinde belirgin bir düşüş gözlemlenmektedir.

Tablo 4.1. Modellerin Performans Ölçütleri

Model	Etiketleme Şeması	Doğruluk	Makro Kesinlik	Makro Duyarlılık	Makro F1	Ağırlıklı F1
BERTurk	non-BIO	0,91	0,93	0,83	0,87	0,91
mBERT	non-BIO	0,93	0,92	0,85	0,87	0,93
XLM-R	non-BIO	0,91	0,90	0,81	0,85	0,91
DistilBERTurk	non-BIO	0,90	0,88	0,78	0,82	0,88
ConvBERTurk	non-BIO	0,91	0,91	0,83	0,86	0,91
BERTurk	BIO	0,84	0,85	0,72	0,72	0,84
mBERT	BIO	0,86	0,88	0,69	0,70	0,87
XLM-R	BIO	0,88	0,89	0,72	0,80	0,87
DistilBERTurk	BIO	0,82	0,84	0,67	0,69	0,82
ConvBERTurk	BIO	0,86	0,87	0,72	0,74	0,86

BERTurk modelinin makro F1 skoru 0,72'ye, DistilBERTurk modelinin ise 0,69'a kadar gerilemiştir. Buna karşın XLM-R modeli, BIO formatta da görece daha güçlü kalmayı başarmış, 0,88 doğruluk ve 0,80 makro F1 skoru ile öne çıkmıştır. ConvBERTurk modeli ise 0,86 doğruluk ve 0,74 makro F1 ile XLM-R modelinin ardından en iyi sonucu veren ikinci sistem olmuştur. Bu durum, XLM-R modelinin hem format farkına hem de sınıf çeşitliliğine karşı daha dirençli bir yapıda olduğunu göstermektedir. mBERT modeli de 0,70

makro F1 değeri ile BIO formatta makul düzeyde kalabilmiştir. Ancak, non-BIO formatta olduğu kadar güçlü bir performans gösterememiştir.

Sınıf bazlı analizler incelendiğinde, non-BIO formatta mBERT modelinin hemen her sınıfta oldukça başarılı olduğu görülmektedir. EK 14'te yer alan verilere göre, mBERT; NUMBER/DATE (0,9671), AWARD (1,0000), SONG/ALBUM (0,9438), PERSON (0,9498) ve SOCIAL_MEDIA (0,9000) gibi kategorilerde yüksek F1 skorları elde etmiştir. Bu başarı, modelin belirgin dil kalıplarını ve semantik yapıları etkili biçimde ayırt edebildiğini göstermektedir. ConvBERTurk modeli de SOCIAL_MEDIA (0,9143), RELATIONSHIP (0,8387) ve EVENT (0,8364) gibi sınıflarda başarı sergilemiştir. XLM-R modeli ise genel dengede kalarak özellikle MOVIE/TV (0,8470) ve HEALTH (0,7887) sınıflarında tatmin edici sonuçlar vermiştir. DistilBERTurk modeli ise özellikle SONG/ALBUM (0,4865) ve ORGANIZATION (0,6759) gibi sınıflarda daha zayıf kalmıştır.

BIO formatındaki sınıf bazlı değerlendirmelere bakıldığında, veri karmaşıklığının modeller üzerindeki etkisi daha net bir biçimde ortaya çıkmaktadır. EK 15'te yer alan bulgular, özellikle I- etiketlerinde F1 skorlarının ciddi ölçüde düştüğünü göstermektedir. I-RELATIONSHIP sınıfında bazı modellerin F1 skoru 0,0000 olarak kalmış, B-HEALTH, B-LOCATION, I-ORGANIZATION gibi etiketlerde ise skorlar sıklıkla 0,60'ın altında kalmıştır. Ancak bu tabloya rağmen, XLM-R modeli BIO formatta da belirli sınıflarda güçlü performans sergilemiştir. Özellikle B-PERSON (0,9477), I-PERSON (0,9137), B-NUMBER/DATE (0,8515), I-NUMBER/DATE (0,9000) ve I-AWARD (0,9429) gibi sınıflarda yüksek başarı göstermiştir. Bu bulgu, XLM-R modelinin özellikle kişi adları, tarih ifadeleri ve ödül türü varlık isimlerinde oldukça etkili olduğunu göstermektedir. ConvBERTurk modeli de B-PERSON (0,9107) ve B-AWARD (0,9333) gibi sınıflarda dikkate değer skorlar üretmiş, fakat bazı alt etiketlerde düşük sonuçlar almıştır. BERTurk modeli ise genel olarak tüm sınıflarda orta düzeyde kalmıştır. DistilBERTurk ise birçok sınıfta beklenenin altında kalarak en düşük genel performansı göstermiştir.

Sonuç olarak, yapılan bu analizler doğrultusunda non-BIO formatta en iyi genel performansı mBERT modeli göstermiş, bunu ConvBERTurk ve XLM-R takip etmiştir. BIO etiketleme formatında ise XLM-R modeli, özellikle kişi adları ve tarih bilgileri gibi sık kullanılan varlık türlerinde en başarılı sistem olarak öne çıkmıştır. Elde edilen bulgular hem model yapılarının hem de etiketleme formatlarının performans üzerinde doğrudan etkili olduğunu ortaya koymaktadır.

Bu tezde ulařılan bulgular, daha nce Trke dilinde NER zerine yapılmıř bazı yksek lisans tezleriyle karřılařtırıldıđında dikkat ekici bir geliřmiřlik dzeyi sergilemektedir. rneđin Khudoyberdieva [28] tarafından yrtlen alıřmada yalnızca BIO etiketleme formatı kullanılmıřtır. Ancak bu yapı zerinde detaylı bir etiketleme sistemi geliřtirilmiřtir. alıřmada dođrudan 23 farklı varlık tr tanımlanmıř ve bu etiketlerle ince taneli bir sınıflandırma yapılmıřtır. Kullanılan etiketler arasında Kurum, Mahkeme, Savcı, Yasa, Ceza, Avukat gibi Trkeye zg kavramlar da yer almaktadır. BERT tabanlı  farklı model (BERTurk, mBERT, XLM-R) kullanılarak genel varlık trlerinde 93,34 ve alt kategoriler dzeyinde ise 79,82 F1 skoru elde edilmiřtir. Ancak alıřmada yalnızca klasik deđerlendirme metrikleri kullanılmıřtır. Etiketleme formatlarının karřılařtırılması yapılmamıř, model kararlarının aıklanabilirliđi ve hata analizi gibi ileri dzey yntemlere yer verilmemiřtir. Bu bađlamda, etiket řeması bakımından kapsamlı olsa da analiz eřitliliđi aısından sınırlı kalmaktadır.

İncidelen [29] tarafından gerekleřtirilen alıřmada ise dnřtrc tabanlı modellerin Trke hukuk metinlerine alan uyarlaması incelenmiřtir. Bu alıřmada yalnızca BIO etiketleme formatı kullanılmıřtır. Etiket sayısı altı ile sınırlı tutulmuřtur. BERTurk ve mBERT'e ek olarak hukuk alanına zg LegalBERT ve ELECTRA gibi modeller de deđerlendirilmiřtir. LegalBERT modeli ile 0,83 makro F1 skoru elde edilmiřtir. Tez, alan uyarlamasına odaklanmış ve hukuk metinlerindeki model bařarımını lmeyi amalamıřtır. Ancak aıklanabilir yapay zek teknikleri, etiketleme řemaları arası karřılařtırmalar veya model karar yapısına dair detaylı analizler bu alıřmada da bulunmamaktadır.

Sonuç olarak, her iki tez de kendi bađlamında nemli katkılar sunmaktadır. Ancak bu tez alıřması, hem etiket sayısı ve eřitliliđi aısından hem de aıklanabilirlik ve analiz derinliđi aısından daha geliřmiř bir yapı sergilemektedir. BIO ve non-BIO etiketleme sistemlerinin birlikte ele alınması, LIME, t-SNE ve hata analizi gibi grsel ve yorumlayıcı tekniklerin sistematik kullanımı, bu alıřmayı mevcut literatrde ne ıkaran gl ynler arasında yer almaktadır.

4.4. LIME Yntemi ile Model Aıklanabilirliđinin Deđerlendirilmesi

Bu blmde, tez kapsamında eđitilen NER modellerinin karar verme srelerine ynelik aıklanabilirlik deđerlendirmeleri sunulmaktadır. Bu amala, Ribeiro ve arkadařları tarafından geliřtirilen LIME (Local Interpretable Model-Agnostic Explanations) yntemi uygulanmıřtır [50]. LIME, herhangi bir sınıflandırıcıdan bađımsız olarak alıřabilen ve

belirli bir örnek çerçevesinde modelin kararlarını yorumlanabilir şekilde açıklamaya yönelik yerel doğrusal modellemeler oluşturan bir tekniktir. Modelin sınıflandırma kararı, orijinal örneğe benzer sentetik girdiler üretilerek bu girdiler üzerinde eğitilen basit bir açıklayıcı model aracılığıyla analiz edilmektedir.

LIME yöntemi, NER bağlamında bir kelimenin belirli bir etiketle eşlenmesinde hangi çevresel sözcüklerin ne düzeyde katkı sunduğunu kantitatif olarak ortaya koyabilmektedir. Böylece yalnızca modelin çıktısı değil, bu çıktının hangi girdilerle üretildiği de ayrıntılı biçimde analiz edilebilmektedir. Bununla birlikte, LIME'ın yerel bir açıklama yöntemi olduğu göz önünde bulundurulmalıdır. Elde edilen açıklamalar yalnızca analiz edilen örneğe özgüdür ve modelin genel davranışlarını temsil etme veya veri kümesi geneline yönelik çıkarımlarda bulunma amacı taşımamaktadır [50].

Analiz kapsamında, her bir model için hem BIO hem de non-BIO etiketleme formatları ile eğitilen sürümleri kullanılarak LIME uygulamaları gerçekleştirilmiştir. Belirli örnek cümleler üzerinden hedef kelimeler seçilmiştir. Bu kelimelere ilişkin sınıflandırma kararına etki eden diğer sözcüklerin katkı düzeyleri hesaplanarak görselleştirilmiştir. Görselleştirmeler aracılığıyla katkı sağlayan veya sınıflandırma üzerinde negatif etki oluşturan sözcükler renk kodlaması ile sunulmuştur. Katkı skorlarının dağılımları ayrıca histogramlar ile desteklenmiştir. Hedef kelimenin vurgulandığı cümle gösterimleri ile bağlam içi analiz güçlendirilmiştir.

Uygulanan LIME analizleri, modelin yalnızca etiketleme başarımını değil, aynı zamanda bağlam algısını, karar mekanizmalarını ve potansiyel sınıflandırma hatalarını da nitelikli şekilde gözleme imkânı sunmaktadır. Bu yönüyle LIME, model çıktılarının yorumlanabilirliğini artıran ve model davranışlarını örnek bazlı açıklamalarla görünür kılan önemli bir analiz aracı olarak değerlendirilmiştir.

4.4.1. non-BIO Etiketleme Üzerine LIME Analizi

Bu bölümde, non-BIO etiketleme formatıyla eğitilmiş olan beş farklı modelin karar verme süreçlerinin açıklanabilirliği, LIME yöntemi aracılığıyla değerlendirilmiştir. LIME, bir modelin belirli bir örnek bağlamındaki sınıflandırma kararlarını yorumlayabilmek amacıyla, sentetik veri üretimi ve lokal doğrusal modelleme uygulayarak etki düzeylerini sayısal olarak ortaya koyar [50]. Uygulanan analizlerde toplam 40 farklı test örneği kullanılmıştır. Her biri gerçek veri setinden seçilmiş cümlelerde yer alan 40 hedef kelime, modellerin etiketleme kararlarına göre değerlendirilmiştir. Bu kelimeler 12 farklı etiket

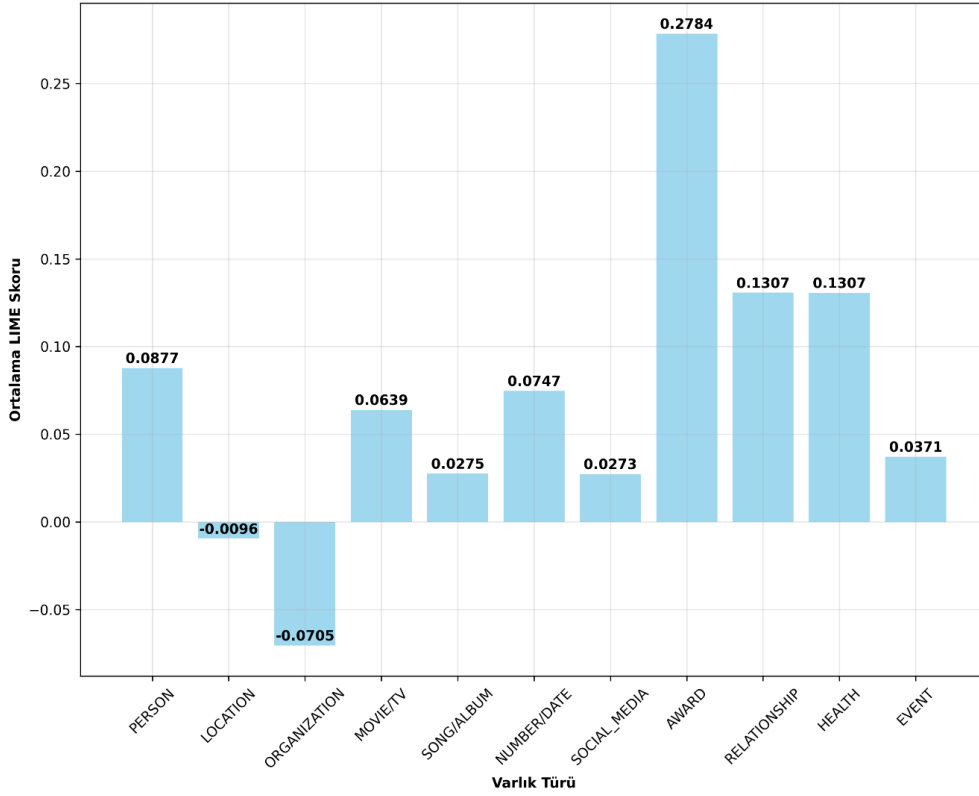
kategorisine (örneğin PERSON, LOCATION, ORGANIZATION, AWARD vb.) dağılmıştır. Her kelime için analiz edilen cümle, LIME katkı skorları ile değerlendirilmiş ve her model için ortalama katkı değeri ile standart sapma hesaplanmıştır.

Sayısal değerlere göre en yüksek ortalama LIME katkı skoru mBERT modelinde gözlemlenmiş olup, bu model 0,1031 ortalama katkı ve 0,0587 standart sapma değerleriyle öne çıkmıştır. Diğer modellerden XLM-R 0,0902, BERTurk 0,0867, ConvBERTurk 0,0779 ve DistilBERTurk ise 0,0634 ortalama katkı değeri sunmuştur. Özellikle mBERT'in düşük varyansla birlikte yüksek katkı skoru üretmesi, bu modelin sınıflandırma kararlarında bağlamsal olarak daha istikrarlı bir performans gösterdiğine işaret etmektedir. İlgili bulgular Tablo 4.2'de sunulmaktadır.

Tablo 4.2. Modellerin LIME Skor Ortalamaları (non-BIO)

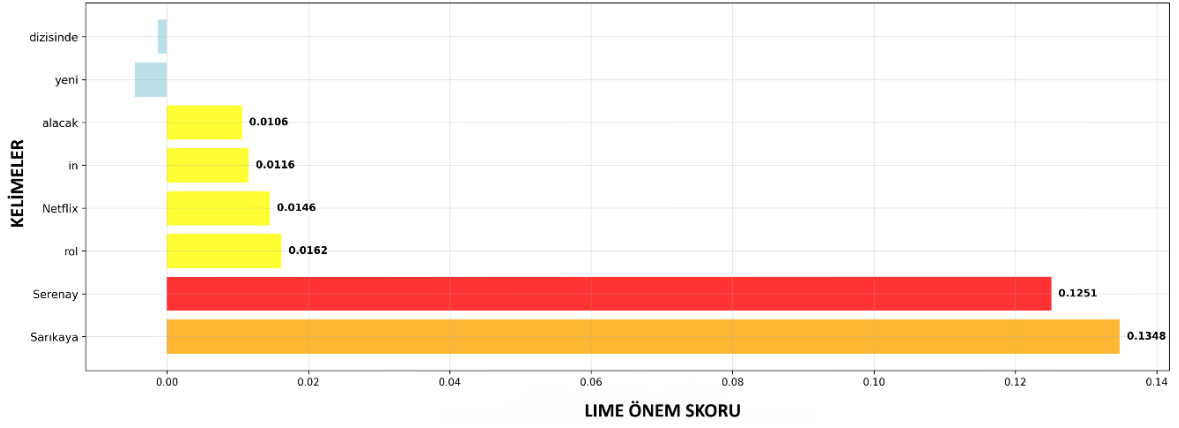
Model	Ortalama Skor	Standart Sapma
mBERT	0,1031	0,0587
XLM-R	0,0902	0,0673
BERTurk	0,0867	0,0747
ConvBERTurk	0,0779	0,0675
DistilBERTurk	0,0634	0,0732

Varlık türlerine göre mBERT modelinin ortalama LIME katkı skorları incelendiğinde; en yüksek katkı AWARD (0,2784), RELATIONSHIP (0,1307) ve HEALTH (0,1307) etiketlerinde hesaplanmıştır. PERSON etiketi için katkı skoru 0,0877 olup, modelin bu sınıfta bağlamsal ipuçlarını güçlü şekilde değerlendirdiği görülmektedir. Diğer yandan ORGANIZATION (-0,0705) ve LOCATION (-0,0096) etiketlerinde negatif katkı değerleri gözlemlenmiştir. Bu durum modelin bu tür varlıklarda karar verirken daha az belirgin veya tutarsız bağlamlardan etkilendiğini göstermektedir. Söz konusu etiket türlerine ilişkin ortalama katkı değerleri Şekil 4.21'de sunulmaktadır.



Şekil 4.21. Etiket Türüne Göre LIME Skorları (mBERT, non-BIO)

PERSON etiketi özelinde yürütülen örnek bir analizde, “Serena Sarıkaya Netflix’in yeni dizisinde rol alacak.” cümlesi üzerinden LIME açıklaması yapılmıştır. Hedef kelime olarak belirlenen “Serena” ifadesi için, en yüksek katkı değeri 0,1348 ile “Sarıkaya” kelimesine ve ikinci en yüksek katkı ise 0,1251 ile “Serena” kelimesine ait olmuştur. Ayrıca “rol” (0,0162), “Netflix” (0,0146), “in” (0,0116) ve “alacak” (0,0106) gibi kelimeler de karar sürecini destekleyen bağlamsal öğeler olarak katkı sağlamıştır. Bu açıklama sonucunda modelin yalnızca hedef kelimenin morfolojik biçimine değil, çevresindeki semantik bağlama da duyarlı olduğu görülmektedir. İlgili çıktı Şekil 4.22’de yer almaktadır.



Şekil 4.22. PERSON Etiketi İçin Örnek LIME Açıklaması (mBERT, non-BIO)

Tüm bu analizler sonucunda, non-BIO etiketleme formatında en yüksek açıklanabilirlik başarımı Google mBERT modeli tarafından sergilenmiştir. Hem ortalama katkı değerinin yüksek olması hem de katkı skorlarının varyansının düşük düzeyde seyretmesi, bu modelin kararlarının hem güçlü hem de tutarlı biçimde bağlam temelli olduğunu göstermektedir. Ancak, LIME analizinin doğası gereği yalnızca lokal düzeyde karar açıklaması sunduğu ve elde edilen bulguların genellenebilirliği konusunda sınırlılık taşıdığı göz önünde bulundurulmalıdır [50].

4.4.2. BIO Etiketleme Üzerine LIME Analizi

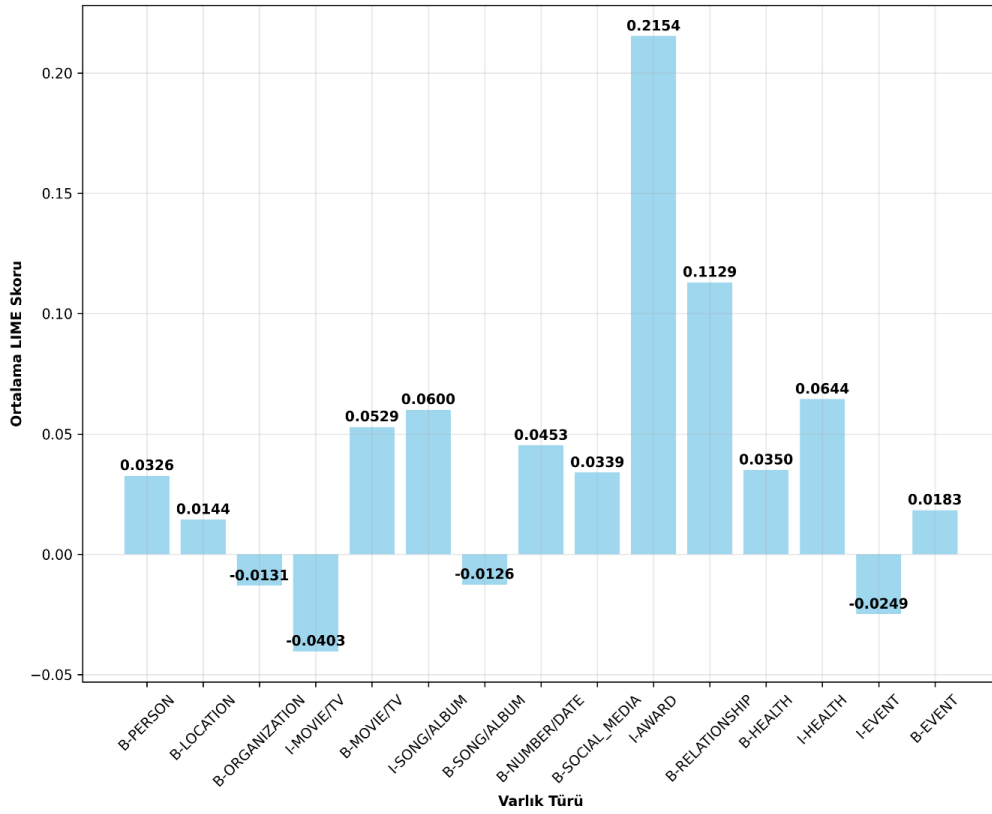
BIO etiketleme formatı ile eğitilen beş modelin sınıflandırma kararlarının açıklanabilirliği, önceki bölümde olduğu gibi LIME yöntemi kullanılarak incelenmiştir. Bu doğrultuda, her model için aynı 40 cümlelik test örnek kümesi kullanılmıştır. 12 farklı etiket türünü temsil eden hedef kelimeler üzerinden analiz gerçekleştirilmiştir.

Analiz sonuçlarına göre, en yüksek ortalama LIME katkı skoruna sahip model 0,1125 ile mBERT olmuştur. Ayrıca bu model, 0,0571 değeriyle en düşük standart sapmaya da sahiptir. Bu durum, modelin sadece yüksek katkı skorları üretmediğini, aynı zamanda bu katkıların örnekler arasında daha tutarlı olduğunu göstermektedir. Diğer modellerin ortalama skorları ise sırasıyla BERTurk (0,1028), XLM-R (0,0923), ConvBERTurk (0,0812) ve DistilBERTurk (0,0696) şeklinde sıralanmaktadır. Bu bağlamda, BIO etiketleme yapısında da en yüksek açıklanabilirlik başarımı mBERT tarafından sergilenmiştir. İlgili karşılaştırmalı sonuçlar Tablo 4.3'te yer almaktadır.

Tablo 4.3. Modellerin LIME Skor Ortalamaları (BIO)

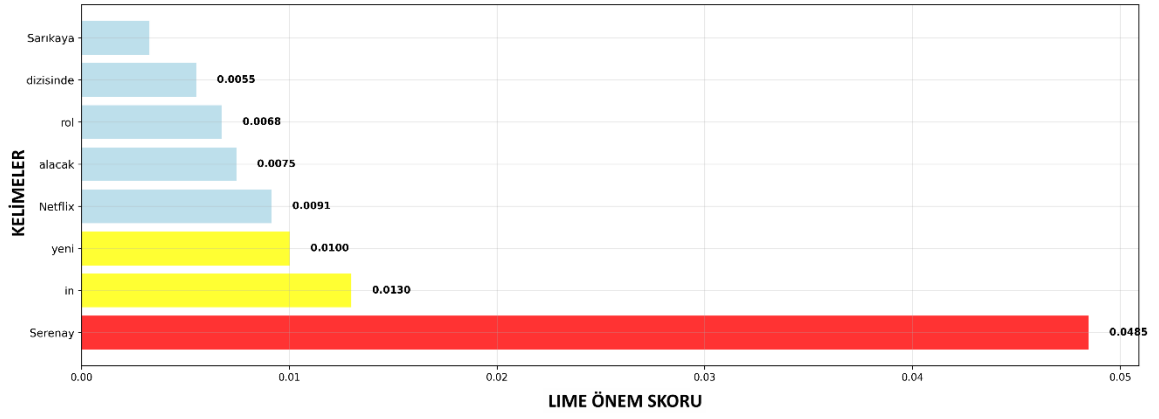
Model	Ortalama Skor	Standart Sapma
mBERT	0,1125	0,0571
BERTurk	0,1028	0,0669
XLM-R	0,0923	0,0712
ConvBERTurk	0,0812	0,0658
DistilBERTurk	0,0696	0,0744

Google mBERT modelinin etiket türlerine göre hesaplanan ortalama katkı skorları Şekil 4.23'te sunulmuştur. Bu görselleştirme, modelin bazı etiket türlerinde bağlamsal farkındalığının daha yüksek olduğunu ortaya koymaktadır. En yüksek katkı değeri AWARD (0,2154) etiketine aittir. Bunu B-RELATIONSHIP (0,1129) ve I-HEALTH (0,0644) etiketleri takip etmektedir. B-PERSON etiketi için katkı değeri 0,0326 olup, modelin bu tür varlıkları sınıflandırırken daha az belirgin bağlamsal sinyallere dayandığı görülmektedir. Negatif katkı değerlerinin en çok gözlemlendiği etiketler arasında B-ORGANIZATION (–0,0403) ve I-EVENT (–0,0249) yer almaktadır. Bu durum, modelin bazı etiket türlerini karıştırma eğiliminde olabileceğini veya bağlamdan bağımsız kararlar verdiğini düşündürmektedir.



Şekil 4.23. Etiket Türüne Göre LIME Skorları (mBERT, BIO)

Şekil 4.24’te, Google mBERT modeli tarafından B-PERSON etiketiyle işaretlenen bir örnek (Serenay Sarıkaya Netflix’in yeni dizisinde rol alacak.) üzerinde yapılan LIME açıklaması sunulmuştur. Bu cümlede hedef olarak belirlenen “Serenay” kelimesi için en yüksek katkı 0,0445 ile yine “Serenay” kelimesine aittir. Bunu sırasıyla “in” (0,0130), “yeni” (0,0100), “Netflix” (0,0091), “alacak” (0,0075), “rol” (0,0068), “dizisinde” (0,0055) ve “Sarıkaya” (0,0045) izlemiştir. Bu katkı dağılımı, modelin kararını sadece hedef kelimenin yüzey biçimine değil, aynı zamanda çevresindeki bağlamsal kelimelere de dayandırdığını göstermektedir. Özellikle soyadın katkı değerinin düşük kalması, modelin ad-soyad bütünlüğünü BIO etiketleme formatında önceki formata (non-BIO) kıyasla daha zayıf şekilde yakaladığını göstermektedir.



Şekil 4.24. B-PERSON Etiketini İçin Örnek LIME Açıklaması (mBERT, BIO)

Yapılan analizler sonucunda, BIO etiketleme formatında da en güçlü açıklanabilirlik kapasitesi Google mBERT modeli tarafından ortaya konmuştur. Hem ortalama katkı değerinin yüksek olması hem de etiket türlerine göre dağılımın dengeli seyretmesi, bu modelin kararlarının bağlamdan beslenen yapıda olduğunu ortaya koymaktadır. Ancak, LIME yönteminin yalnızca lokal açıklamalar sunduğu ve modelin genel davranışını tümüyle temsil edemeyeceği dikkate alınmalıdır [50].

5. SONUÇ VE ÖNERİLER

5.1. Sonuç

Bu tez çalışmasında, Türkçe magazin haberlerinden oluşturulan özel bir veri kümesi üzerinde NER sistemlerinin başarımı, farklı etiketleme formatları ve derin öğrenme tabanlı modeller kullanılarak kapsamlı bir biçimde incelenmiştir. Çalışmanın temel hedeflerinden biri, BIO ve non-BIO etiketleme şemalarının model performansına etkisini karşılaştırmalı olarak değerlendirmek olmuştur. Aynı zamanda mBERT, XLM-R, BERTurk, DistilBERTurk ve ConvBERTurk olmak üzere beş farklı ön-eğitilmiş dönüştürücü tabanlı modeller ile deneyler gerçekleştirilmiş, bu modellerin hem doğruluk temelli performansları hem de açıklanabilirlik düzeyleri ayrı ayrı analiz edilmiştir.

Elde edilen sonuçlar, etiketleme formatlarının model başarımı üzerinde belirgin farklar oluşturduğunu ortaya koymuştur. Genel olarak non-BIO formatında, daha yüksek doğruluk ve F1 skorları elde edilmiştir. Bunun en temel nedeni, non-BIO etiketleme sisteminin etiket geçişlerini basitleştirerek sınıflandırma görevini daha doğrudan hâle getirmesidir. Öte yandan BIO format, çok kelimeli varlıkların ayrıştırılmasına olanak tanınması bakımından yapısal avantaj sunmakta ve modelin bağlam duyarlılığını daha fazla zorlamaktadır. Bu bağlamda, veri kümesinin yapısına ve kullanım amacına bağlı olarak her iki formatın da tercih edilebilir olduğu görülmüştür.

Model karşılaştırmaları incelendiğinde hem BIO hem de non-BIO etiketleme sistemlerinde en yüksek başarı Google mBERT modeli tarafından elde edilmiştir. mBERT modeli hem makro hem de ağırlıklı F1 skorlarında en yüksek değerlere ulaşmış ve çoğu sınıf için dengeli sonuçlar üretmiştir. Bu başarı, modelin çok dilli yapısı, geniş çaplı ön eğitim verisi ve yapısal esnekliği ile ilişkilendirilmiştir. XLM-R ve BERTurk modelleri de rekabetçi performanslar sergilemiştir. ConvBERTurk ve DistilBERTurk modelleri ise daha hafif mimariye sahip oldukları için genel olarak daha düşük doğruluk düzeylerinde kalmış ancak bazı sınıflarda anlamlı sonuçlar üretmiştir.

Çalışmada yalnızca genel metrikler değil, aynı zamanda sınıf bazlı analizler ve görsel değerlendirmeler de yapılmıştır. t-SNE kümelenme görselleri ve normalize edilmiş karışıklık matrisleri üzerinden gerçekleştirilen incelemeler, modellerin farklı etiket türleri arasındaki ayrım gücü ve hata eğilimlerini ortaya koymuştur. Bu değerlendirmeler, özellikle mBERT modelinin bağlamdan beslenen bir karar yapısı geliştirdiğini ve anlam bakımından benzer sınıflar arasında görece daha az hata yaptığını göstermektedir.

Bununla birlikte, çalışmada açıklanabilir yapay zekâ yöntemlerinden biri olan LIME tekniği uygulanarak, modellerin belirli bir sınıflandırma kararını hangi kelimelere dayandırdığı analiz edilmiştir. LIME sonuçları, en yüksek katkı skorlarının genellikle mBERT modelinde gözlemlendiğini ve bu modelin kararlarını daha belirgin bağlamsal işaretlere dayandırdığını ortaya koymuştur. Formatlar karşılaştırıldığında, non-BIO formatında elde edilen katkı skorları genellikle daha yüksek olmuştur. Bu durum da modelin daha sade etiket yapısıyla daha net bağlamsal ayrımlar yapabildiğini göstermiştir. BIO formatta ise daha ayrıntılı ve parçalı kararların üretildiği, ancak bazı etiket geçişlerinde karışıklıkların arttığı görülmüştür. LIME analizlerinin yalnızca lokal karar bölgelerine ışık tuttuğu ve modelin bütünsel davranışını temsil etmekte sınırlı olduğu da unutulmamalıdır.

Sonuç olarak, bu tez çalışmasında etiketleme şeması seçiminin, model mimarisinin ve açıklanabilirlik analizlerinin NER sistemleri üzerindeki etkilerini çok boyutlu olarak ortaya koymuştur. Hem doğruluk hem de karar gerekçelerinin şeffaflığı açısından yapılan karşılaştırmalar, ilerleyen çalışmalarda hem uygulama geliştirme hem de yöntemsel iyileştirme açısından yol gösterici niteliktedir.

5.2. Öneriler

Bu çalışmanın bulgularına dayanarak gelecekte yapılacak araştırmalar için bazı öneriler sunulabilir. Etiketleme şemalarının model başarımını önemli ölçüde etkilediği görülmüştür. Bu nedenle veri kümesinin yapısına uygun olarak BIO ya da non-BIO formatı tercih edilmelidir. Çok kelimededen oluşan varlıkların bulunduğu görevlerde BIO formatı daha faydalı olabilir. Daha basit yapıları görevlerde ise non-BIO formatı daha dengeli sonuçlar verebilir. Bunun yanında yalnızca bu iki formatla sınırlı kalmayıp, IOBES gibi farklı etiketleme sistemlerinin de incelenmesi önerilmektedir. Böylece farklı veri türlerinde hangi sistemin daha başarılı olduğu daha net anlaşılabilir. Model seçimi yapılırken yüksek doğruluk ve güvenilirlik hedefleniyorsa çok dilli büyük modeller kullanılabilir. Özellikle mBERT ve XLM-R gibi modeller güçlü sonuçlar vermektedir. Daha sınırlı kaynaklarla çalışılması gereken durumlarda ise DistilBERTTurk gibi küçük modeller tercih edilebilir. Açıklanabilirlik açısından LIME yöntemi, modelin karar verirken hangi kelimelere dayandığını anlamayı sağlamaktadır. Bu yöntem etkili olsa da ileride SHAP gibi başka açıklama yöntemlerinin de test edilmesi faydalı olacaktır. Ayrıca bu çalışma sadece magazin haberleri ile sınırlı kalmıştır. Gelecekte sosyal medya, haber, finans gibi farklı alanlardaki metinlerle de deneyler yapılması, modellerin genellenebilirliğini arttıracaktır. Son olarak,

varlık ismi tanıma sistemlerinde yalnızca doğru sonuç üretmek değil, bu sonuçların gerekçelerini anlaşılır biçimde sunabilmek de önem taşımaktadır. Bu doğrultuda, açıklanabilirlik odaklı yöntemlerin sistemin farklı aşamalarına entegre edilmesi, yapay zekâ uygulamalarının daha yorumlanabilir biçimde çalışmasına katkı sağlayabilir.

KAYNAKLAR

- [1] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification,” *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.
- [2] E. F. Tjong Kim Sang and F. De Meulder, “Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition,” in *Proc. 7th Conf. on Natural Language Learning at HLT-NAACL*, 2003, pp. 142–147.
- [3] R. Yeniterzi, “Exploiting morphology in Turkish named entity recognition,” in *Proc. ACL Student Research Workshop*, 2011.
- [4] G. Çelikkaya, D. Torunoğlu, and G. Eryiğit, “Named entity recognition on real data: A preliminary investigation for Turkish,” in *Proc. AICT*, 2013, pp. 1–5.
- [5] D. Küçük, G. Jacquet, and R. Steinberger, “Named entity recognition on Turkish tweets,” in *Proc. LREC*, 2014.
- [6] S. Özkaya and B. Diri, “Named entity recognition by conditional random fields from Turkish informal texts,” in *Proc. SIU*, 2011, pp. 662–665.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186. Available: <https://arxiv.org/abs/1810.04805>
- [8] S. Cucerzan and D. Yarowsky, “Language independent named entity recognition combining morphological and contextual evidence,” in *Proc. Joint SIGDAT Conf. on EMNLP and VLC*, 1999, pp. 90–99.
- [9] Ö. Bayraktar, *Person name recognition in Turkish financial texts by using local grammar approach*, MSc thesis, Middle East Technical University, 2007.
- [10] S. R. Yavuz, *Named entity recognition in Turkish with Bayesian learning and hybrid approaches*, MSc thesis, Middle East Technical University, 2011.
- [11] S. Özkaya, *Türkçe metinlerde şartlı rastgele alanlarla varlık ismi tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, 2013.

- [12] B. Eken, *Kısa metinlerde varlık ismi tanıma*, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, 2015.2015.
- [13] O. Öztürkmenoğlu, *Extraction of named entities from Turkish document collections*, MSc thesis, Dokuz Eylül University, 2018.
- [14] E. K. Akkaya, *Deep neural networks for named entity recognition on social media*, MSc thesis, Hacettepe University, 2018.
- [15] A. Güneş, *Türkçe'de varlık ismi tanıma*, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, 2018.
- [16] M. Keskin, *Otomatik veri etiketleme ile varlık ismi tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, 2019.
- [17] Y. Eşref, *Türkçe dizi etiketleme için sinir ağ modelleri*, Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2019.
- [18] Ö. Çoban, *Attribute inference over real-world online social networks: A comprehensive privacy analysis*, MSc thesis, Çukurova University, 2021.
- [19] D. Dinç, *Financial named entity recognition for Turkish news texts*, MSc thesis, Middle East Technical University, 2022.
- [20] M. Gencer, *Yapısal olmayan metinler için adlandırılmış varlık tanıma algoritmaları ve uygulamaları*, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, 2022.
- [21] B. Çarık, *Enhancing named entity recognition in Turkish by integrating external knowledge and extra layers into transformer-based models*, MSc thesis, Sabancı University, 2022.
- [22] K. Ç. Gürcan, *Named entity recognition with neural networks and pretrained word embeddings*, MSc thesis, Çankaya University, 2023.
- [23] E. Altıntaş, *Dile özgü öznitelikleri kullanan bir varlık ismi tanıma sistemi geliştirilmesi*, Yüksek Lisans Tezi, Ege Üniversitesi, 2023.

- [24] Y. P. Kılıç, *Named entity recognition and explainability analysis on Turkish sports news texts*, MSc thesis, Middle East Technical University, 2023.
- [25] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed., Draft, Jan. 12, 2025. Available: [https:// web.stanford. edu/~jurafsky/slp3](https://web.stanford.edu/~jurafsky/slp3)
- [26] V. Yadav and S. Bethard, “A survey on recent advances in named entity recognition from deep learning models,” in *Proc. 27th Int. Conf. Computational Linguistics (COLING)*, 2018, pp. 2145–2158.
- [27] S. Warty et al., “Systematic literature review on named entity recognition: Approach, method, and application,” *Stat., Optim. Inf. Comput.*, vol. 12, pp. 907–942, Jul. 2024.
- [28] L. Khudoyberdieva, *Türkçe alt kategorili varlık ismi tanıma ile otomatik etiketleme*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, 2024.
- [29] M. İncidelen, *Düşük kaynaklı dillerde transformer tabanlı dil modellerinin alan adaptasyonu: Türkçe hukuk metinleri üzerine bir çalışma*, Yüksek Lisans Tezi, Fırat Üniversitesi, 2025.
- [30] N. Chinchor, “MUC-6 Named Entity Task Definition,” in *Proc. 6th Message Understanding Conf. (MUC-6)*, Columbia, MD, USA, 1995.
- [31] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O’Reilly Media, 2009.
- [32] A. Vaswani et al., “Attention is All You Need,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.
- [33] DBMDZ, “dbmdz/bert-base-turkish-cased,” Hugging Face. Available: [https:// huggingface. co/dbmdz/bert-base-turkish-cased](https://huggingface.co/dbmdz/bert-base-turkish-cased)
- [34] J. Rogers, O. Kovaleva, and A. Rumshisky, “A Primer in BERTology: What We Know About How BERT Works,” *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, 2020.

- [35] DBMDZ, “BERTurk Model Description,” GitHub. Available: <https://github.com/dbmdz/berts>
- [36] Hugging Face, “google-bert/bert-base-multilingual-cased.” Available: <https://huggingface.co/google-bert/bert-base-multilingual-cased>
- [37] T. Pires, E. Schlinger, and D. Garrette, “How multilingual is Multilingual BERT?,” in *Proc. 57th Annu. Meet. Assoc. Comput. Linguist.*, 2019, pp. 4996–5001.
- [38] A. Conneau et al., “Unsupervised Cross-lingual Representation Learning at Scale,” in *Proc. 58th Annu. Meet. Assoc. Comput. Linguist.*, 2020, pp. 8440–8451.
- [39] Facebook AI, “xlm-roberta-base,” Hugging Face. Available: <https://huggingface.co/xlm-roberta-base>
- [40] T. Kudo and J. Richardson, “SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing,” in *Proc. 2018 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations*, 2018, pp. 66–71.
- [41] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv preprint, arXiv:1907.11692, 2019.
- [42] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter,” arXiv preprint, arXiv:1910.01108, 2019.
- [43] S. Schweter, “Turkish DistilBERT Model,” Hugging Face Model Hub. Available: <https://huggingface.co/dbmdz/distilbert-base-turkish-cased>
- [44] G. Hinton, O. Vinyals, and J. Dean, “Distilling the Knowledge in a Neural Network,” arXiv preprint, arXiv:1503.02531, 2015.
- [45] C. Bucilă, R. Caruana, and A. Niculescu-Mizil, “Model Compression,” in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2006, pp. 535–541.

- [46] A. Rogers, O. Kovaleva, and A. Rumshisky, “A Primer on Neural Network Models for Natural Language Processing,” *J. Artif. Intell. Res.*, vol. 57, pp. 615–731, 2020.
- [47] Z. Jiang et al., “ConvBERT: Improving BERT with Span-based Dynamic Convolution,” in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 12837–12848, 2020.
- [48] S. Schweter, “Turkish ConvBERT Model,” Hugging Face Model Hub. Available: <https://huggingface.co/dbmdz/convbert-base-turkish-mc4-uncased>
- [49] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018. Available: <https://doi.org/10.1109/ACCESS.2018.2870052>
- [50] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?: Explaining the Predictions of Any Classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 1135–1144. Available: <https://doi.org/10.1145/2939672.2939778>
- [51] Y. Sasaki, “The Truth of the F-measure,” *Teach. Tutor. Mater.*, vol. 1, no. 5, pp. 1–5, 2007.
- [52] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge Univ. Press, 2008.
- [53] J. J. Opitz, “A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice,” *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 1234–1256, 2024.
- [54] D. M. W. Powers, “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [55] T. Saito and M. Rehmsmeier, “The Precision-Recall Plot is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” *PLOS ONE*, vol. 10, no. 3, p. e0118432, 2015. Available: <https://doi.org/10.1371/journal.pone.0118432>

- [56] L. van der Maaten and G. Hinton, “Visualizing Data Using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [57] A. Güven, “Turkish News Dataset,” Hugging Face. Available: https://huggingface.co/datasets/nilguven/turkish_news_dataset
- [58] B. M. Hussein and S. M. Shareef, *An Empirical Study on the Correlation Between Early Stopping Patience and Epochs in Deep Learning*, 2024.
- [59] Council of Higher Education of Turkey (YÖK), “National Thesis Center,” Available: <https://tez.yok.gov.tr>

EKLER

EK 1: Varlık İsim Tanıma Tezleri Örnekleri

Kaynak: Yükseköğretim Kurulu Başkanlığı Ulusal Tez Merkezi [59].

Yazar	Yıl	Tez Adı	Tez Türü
Özkan Bayraktar	2007	Person name recognition in Turkish financial texts by using local grammar approach	Yüksek Lisans
Sermet Reha Yavuz	2011	Named entity recognition in Turkish with Bayesian learning and hybrid approaches	Yüksek Lisans
Serap Özkaya	2013	Türkçe metinlerde şartlı rastgele alanlarla varlık ismi tanıma	Yüksek Lisans
Beyza Eken	2015	Kısa metinlerde varlık ismi tanıma	Yüksek Lisans
Okan Öztürkmenoğlu	2018	Extraction of named entities from Turkish document collections	Doktora
Emre Kağan Akkaya	2018	Deep neural networks for named entity recognition on social media	Yüksek Lisans
Asım Güneş	2018	Türkçe'de varlık ismi tanıma	Yüksek Lisans
Yasin Eşref	2019	Türkçe dizi etiketleme için sinir ağ modelleri	Yüksek Lisans
Mustafa Keskin	2019	Otomatik veri etiketleme ile varlık ismi tanıma	Yüksek Lisans
Önder Çoban	2021	Attribute inference over real-world online social networks: a comprehensive privacy analysis	Doktora
Duygu Dinç	2022	Financial named entity recognition for Turkish news texts	Yüksek Lisans
Mustafa Gencer	2022	Yapısal olmayan metinler için adlandırılmış varlık tanıma algoritmaları ve uygulamaları	Yüksek Lisans

Yazar	Yıl	Tez Adı	Tez Türü
Buse Çarık	2022	Enhancing named entity recognition in Turkish by integrating external knowledge and extra layers into transformers	Yüksek Lisans
Kartal Çağlar Gürcan	2023	Named entity recognition with neural networks and pretrained word embeddings	Yüksek Lisans
Ergin Altıntaş	2023	Dile özgü öznitelikleri kullanan bir varlık ismi tanıma sistemi geliştirilmesi	Doktora
Yüksel Pelin Kılıç	2023	Named entity recognition and explainability analysis on Turkish sports news texts	Yüksek Lisans
Lola Khudoyberdieva	2024	Türkçe alt kategorili varlık ismi tanıma ile otomatik etiketleme	Yüksek Lisans
Mert İncidelen	2025	Düşük kaynaklı dillerde transformer tabanlı dil modellerinin alan adaptasyonu	Yüksek Lisans

EK 2: Modellerin Karşılaştırmalı Özellikleri

Özellik	BERTurk	mBERT	XLM-RoBERTa	DistilBERTurk	ConvBERTurk
Geliştirici Kurum	DBMDZ	Google	Facebook AI	DBMDZ	DBMDZ
Model Mimarisi	BERT	BERT	RoBERTa	BERT (bilgi damıtma tekniğiyle)	ConvBERT
Katman Sayısı	12	12	12	6	12
Parametre Sayısı	Yaklaşık 110 milyon	Yaklaşık 110 milyon	Yaklaşık 125 milyon	Yaklaşık 66 milyon	Yaklaşık 110 milyon
Gömme Katmanları	Simge, konum ve segment	Simge, konum ve segment	Simge ve konum	Simge ve konum	Simge, konum ve segment
Alt Birimleme Yöntemi	WordPiece (32.000)	WordPiece (110.000)	SentencePiece (250.000)	WordPiece (32.000)	WordPiece (32.000)
Segment Gömmesi	Evet	Evet	Hayır	Hayır	Evet
Büyük/Küçük Harf Duyarlılığı	Duyarlı	Duyarlı	Duyarlı	Duyarlı	Duyarsız
Türkçeye Özgülük	Evet	Hayır	Hayır	Evet	Evet
Eğitim Verisi	TR Wikipedia, OSCAR, OPUS, Oflazer	104 dil Wikipedia	CommonCrawl + Wikipedia (2,5 TB)	BERTurk'ten türetilmiş	Türkçe Multilingual C4 (242 GB)

Özellik	BERTurk	mBERT	XLM-RoBERTa	DistilBERTurk	ConvBERTurk
Eğitim Yöntemi	MLM + NSP	MLM + NSP	Yalnızca MLM (dinamik)	MLM + bilgi damıtma	MLM
Model Boyutu (Disk)	Yaklaşık 420 MB	Yaklaşık 660 MB	Yaklaşık 550 MB	Yaklaşık 250 MB	Yaklaşık 420 MB
Görelî Eğitim Süresi	Orta	Uzun	Uzun	Kısa	Orta
Dönüştürücü Yapı Özeti	Standart	Standart	Optimize edilmiş yapı	Azaltılmış (6 katmanlı)	Karma dikkat (öz-dikkat + evrişim)
Yayın Yılı	2020	2018	2020	2021	2021

EK 3: Etiketleme Formatlarına Göre Sınıf Dağılımı

Etiket (BIO)	Eğitim + Doğrulama	Final Test	Etiket (non-BIO)	Eğitim + Doğrulama	Final Test
O	5728	1448	O	5728	1448
B-PERSON	828	189	PERSON	1560	357
I-PERSON	732	168			
B-LOCATION	273	47	LOCATION	397	80
I-LOCATION	124	33			
B-NUMBER/DATE	312	61	NUMBER/DATE	597	112
I-NUMBER/DATE	285	51			
B-ORGANIZATION	126	35	ORGANIZATION	215	58
I-ORGANIZATION	89	23			
B-MOVIE/TV	163	47	MOVIE/TV	339	86
I-MOVIE/TV	176	39			
B-SONG/ALBUM	40	12	SONG/ALBUM	77	23
I-SONG/ALBUM	37	11			
B-EVENT	86	13	EVENT	207	26
I-EVENT	115	13			
B-SOCIAL_MEDIA	57	16	SOCIAL_MEDIA	89	24
I-SOCIAL_MEDIA	32	8			
B-AWARD	41	15	AWARD	146	58
I-AWARD	105	43			
B-RELATIONSHIP	194	37	RELATIONSHIP	207	48
I-RELATIONSHIP	13	11			
B-HEALTH	30	15	HEALTH	62	37
I-HEALTH	32	22			

EK 4: BERTurk Eğitim Sonuçları (non-BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	4	0,8580	0,9085	9,3
8	2	6	0,8609	0,9072	11,1
8	3	5	0,8564	0,9065	9,8
8	4	7	0,8720	0,9114	8,5
8	5	6	0,8552	0,9063	7,3
16	1	9	0,8728	0,9084	9,6
16	2	7	0,8505	0,9050	8,9
16	3	10	0,8540	0,9038	11,2
16	4	8	0,8630	0,9066	10,1
16	5	7	0,8538	0,9063	8,9
32	1	10	0,8513	0,9026	13,1
32	2	10	0,8536	0,9047	10,5
32	3	12	0,8620	0,9063	12,6
32	4	11	0,8556	0,9083	11,4
32	5	12	0,8593	0,9086	12,4

EK 5: mBERT Eğitim Sonuçları (non-BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	7	0,8781	0,9237	18,4
8	2	5	0,8590	0,9207	12,9
8	3	6	0,8352	0,9106	10,9
8	4	7	0,8743	0,9260	10,7
8	5	5	0,8386	0,9132	11,0
16	1	8	0,8452	0,9138	13,1
16	2	8	0,8456	0,9148	12,9
16	3	8	0,8506	0,9181	14,3
16	4	8	0,8510	0,9231	12,8
16	5	10	0,8633	0,9191	16,2
32	1	8	0,8490	0,9168	15,1
32	2	9	0,8531	0,9209	19,6
32	3	11	0,8617	0,9214	18,7
32	4	10	0,8523	0,9171	15,0
32	5	11	0,8666	0,9209	19,4

EK 6: XLM-R Eğitim Sonuçları (non-BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	5	0,7548	0,8962	12,0
8	2	9	0,8441	0,9163	17,6
8	3	7	0,8528	0,9145	17,4
8	4	7	0,8242	0,9109	13,9
8	5	9	0,8438	0,9100	17,4
16	1	9	0,8429	0,9113	15,8
16	2	11	0,8448	0,9124	16,1
16	3	10	0,8211	0,9070	16,1
16	4	11	0,8512	0,9133	17,5
16	5	9	0,8344	0,9132	17,5
32	1	13	0,8105	0,9059	16,7
32	2	17	0,8520	0,9096	22,4
32	3	15	0,8406	0,9126	19,9
32	4	12	0,8463	0,9128	20,0
32	5	13	0,8034	0,9034	17,0

EK 7: DistilBERTurk Eğitim Sonuçları (non-BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	9	0,7914	0,8797	5,6
8	2	7	0,7826	0,8830	5,7
8	3	9	0,7886	0,8775	5,8
8	4	11	0,8202	0,8775	8,3
8	5	9	0,7921	0,8828	6,4
16	1	12	0,7725	0,8706	7,6
16	2	13	0,7918	0,8782	7,4
16	3	12	0,7813	0,8798	6,8
16	4	14	0,7831	0,8828	7,9
16	5	15	0,7988	0,8860	8,5
32	1	20	0,7855	0,8823	11,6
32	2	20	0,7872	0,8777	12,1
32	3	19	0,7650	0,8760	10,1
32	4	20	0,7842	0,8829	11,5
32	5	20	0,7725	0,8834	10,4

EK 8: ConvBERTurk Eğitim Sonuçları (non-BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	8	0,8594	0,9046	11,9
8	2	9	0,8344	0,9022	15,3
8	3	8	0,8261	0,8984	11,9
8	4	9	0,8307	0,9016	15,2
8	5	9	0,8559	0,9059	14,3
16	1	8	0,7826	0,8945	11,2
16	2	10	0,7907	0,8921	12,7
16	3	12	0,8428	0,9040	13,8
16	4	10	0,8401	0,9046	12,6
16	5	10	0,8358	0,8994	14,9
32	1	17	0,8522	0,9035	18,8
32	2	12	0,8215	0,8947	16,2
32	3	12	0,8173	0,8977	17,3
32	4	17	0,8082	0,8947	19,2
32	5	16	0,8452	0,9022	17,8

EK 9: BERTurk Eğitim Sonuçları (BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	7	0,7130	0,8420	10,3
8	2	8	0,7062	0,8297	13,2
8	3	8	0,7016	0,8404	10,1
8	4	7	0,7037	0,8389	8,8
8	5	8	0,7168	0,8328	10,1
16	1	10	0,7040	0,8349	10,9
16	2	10	0,6928	0,8425	11,5
16	3	10	0,7012	0,8390	11,6
16	4	9	0,6968	0,8315	12,6
16	5	9	0,7194	0,8388	11,4
32	1	11	0,6977	0,8377	14,3
32	2	15	0,6935	0,8376	17,3
32	3	13	0,6940	0,8323	13,9
32	4	13	0,6919	0,8363	13,8
32	5	12	0,6893	0,8355	13,8

EK 10: mBERT Eğitim Sonuçları (BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	7	0,6433	0,8460	17,0
8	2	7	0,6345	0,8353	12,7
8	3	6	0,7028	0,8682	16,5
8	4	6	0,6601	0,8628	16,2
8	5	7	0,6834	0,8543	18,4
16	1	9	0,6728	0,8580	18,2
16	2	12	0,6664	0,8580	19,7
16	3	11	0,6561	0,8553	17,5
16	4	10	0,6510	0,8631	17,7
16	5	9	0,6408	0,8455	19,2
32	1	10	0,6616	0,8562	19,6
32	2	12	0,6549	0,8459	21,2
32	3	11	0,6535	0,8634	20,8
32	4	15	0,6691	0,8611	22,7
32	5	13	0,6910	0,8698	23,9

EK 11: XLM-R Eğitim Sonuçları (BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	8	0,7195	0,8790	22,2
8	2	10	0,6986	0,8632	20,7
8	3	8	0,6127	0,8536	17,3
8	4	10	0,7502	0,8910	20,9
8	5	10	0,6812	0,8672	18,9
16	1	11	0,6355	0,8641	17,1
16	2	12	0,6507	0,8669	18,9
16	3	16	0,6739	0,8613	23,3
16	4	13	0,7153	0,8802	21,8
16	5	13	0,6429	0,8610	18,8
32	1	15	0,7135	0,8799	25,7
32	2	21	0,6928	0,8723	29,1
32	3	20	0,6971	0,8731	26,6
32	4	24	0,7965	0,8730	31,2
32	5	18	0,6966	0,8747	25,9

EK 12: DistilBERTurk Eğitim Sonuçları (BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	14	0,6582	0,8165	10,7
8	2	12	0,6471	0,8195	10,5
8	3	11	0,5974	0,8094	7,8
8	4	13	0,6646	0,8208	9,6
8	5	11	0,6879	0,8113	8,9
16	1	14	0,6552	0,8221	8,3
16	2	14	0,6100	0,8073	8,1
16	3	16	0,6452	0,8118	10,9
16	4	15	0,6324	0,8149	16,1
16	5	17	0,6425	0,8053	92,8
32	1	20	0,5999	0,8095	69,8
32	2	23	0,6418	0,7990	70,9
32	3	25	0,6027	0,7959	73,5
32	4	22	0,6152	0,8008	45,0
32	5	24	0,6383	0,8147	13,1

EK 13: ConvBERTurk Eğitim Sonuçları (BIO)

Toplu İşlem Boyutu	Kat	En İyi Döngü	Test Makro F1	Test Ağırlıklı F1	Eğitim Süresi (dk)
8	1	9	0,7034	0,8435	16,8
8	2	10	0,7406	0,8640	17,5
8	3	11	0,7353	0,8436	17,4
8	4	15	0,7405	0,8612	19,5
8	5	12	0,7282	0,8446	15,5
16	1	12	0,6962	0,8550	16,9
16	2	16	0,7219	0,8583	22,1
16	3	18	0,7216	0,8603	21,1
16	4	17	0,7171	0,8501	23,2
16	5	17	0,7094	0,8550	19,7
32	1	25	0,7164	0,8623	26,1
32	2	24	0,6752	0,8312	27,4
32	3	20	0,6828	0,8561	24,6
32	4	23	0,6794	0,8474	29,3
32	5	19	0,7198	0,8674	28,4

EK 14: non-BIO Etiketleme Formatında Modellerin Sınıf Bazlı F1 Skorları

Sınıf	BERTurk	mBERT	XLM-R	DistilBERTurk	ConvBERTurk
O	0,9288	0,9455	0,9354	0,9206	0,9261
PERSON	0,9390	0,9498	0,9447	0,9277	0,9356
LOCATION	0,8103	0,8404	0,8232	0,7468	0,7675
NUMBER/DATE	0,9430	0,9671	0,9588	0,9412	0,9583
ORGANIZATION	0,7763	0,7862	0,8000	0,6759	0,7445
MOVIE/TV	0,8824	0,8974	0,8470	0,7431	0,7982
SONG/ALBUM	0,9434	0,9438	0,8197	0,4865	0,7451
EVENT	0,8056	0,7826	0,7674	0,7778	0,8364
SOCIAL_MEDIA	0,8649	0,9000	0,8718	0,8000	0,9143
AWARD	1,0000	1,0000	0,9647	0,8889	0,9667
RELATIONSHIP	0,7708	0,7839	0,6929	0,6966	0,8387
HEALTH	0,8000	0,6947	0,7887	0,7925	0,7111

EK 15: BIO Etiketleme Formatında Modellerin Sınıf Bazlı F1 Skorları

Sınıf	BERTurk	mBERT	XLN-R	DistilBERTurk	ConvBERTurk
O	0,9251	0,9409	0,9363	0,9226	0,9257
B-PERSON	0,7672	0,8375	0,9477	0,7716	0,9107
I-PERSON	0,7917	0,8453	0,9137	0,7923	0,8720
B-LOCATION	0,5486	0,7581	0,6520	0,4971	0,6127
I-LOCATION	0,5468	0,6875	0,5967	0,5037	0,5185
B-NUMBER/DATE	0,8400	0,8372	0,8515	0,8723	0,8515
I-NUMBER/DATE	0,9167	0,8605	0,9000	0,9032	0,9053
B-ORGANIZATION	0,5679	0,5500	0,5854	0,4810	0,6053
I-ORGANIZATION	0,5385	0,6133	0,5067	0,5135	0,5507
B-MOVIE/TV	0,7805	0,6126	0,6383	0,7209	0,6757
I-MOVIE/TV	0,6950	0,6667	0,7033	0,6753	0,6457
B-SONG/ALBUM	0,6667	0,5000	0,6452	0,5833	0,6957
I-SONG/ALBUM	0,7500	0,6275	0,5500	0,5385	0,6429
B-EVENT	0,5263	0,4667	0,4615	0,5714	0,6667
I-EVENT	0,8571	0,7467	0,7273	0,8148	0,8205
B-SOCIAL_MEDIA	0,7500	0,7273	0,7826	0,7500	0,8333
I-SOCIAL_MEDIA	0,9091	0,7826	0,8000	0,9091	0,8333
B-AWARD	0,8235	0,7000	0,7143	0,7143	0,9333
I-AWARD	0,9600	0,9315	0,9429	0,8980	0,9130
B-RELATIONSHIP	0,7216	0,7293	0,6718	0,6500	0,7442
I-RELATIONSHIP	0,0000	0,0000	0,2857	0,0000	0,5455
B-HEALTH	0,5000	0,3243	0,6061	0,5600	0,6364
I-HEALTH	0,6452	0,6429	0,6000	0,6429	0,6923