

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAPAY ZEKA VE ROBOTİK ANABİLİM DALI

**DUYGU ANALİZİNDE TRANSFORMER TABANLI MODELLERİN
KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

Hayrullah TEMEL
Danışman: Dr. Öğr. Üyesi Erol KINA

VAN – 2025

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAPAY ZEKA VE ROBOTİK ANABİLİM DALI

**DUYGU ANALİZİNDE TRANSFORMER TABANLI MODELLERİN
KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

Hayrullah TEMEL

Tez Savunma Sınavı Jüri Üyeleri

Dr. Öğr. Üyesi İsa AVCI (Başkan)

Dr. Öğr. Üyesi Erol KINA (Danışman)

Dr. Öğr. Üyesi Murat KOCA (Üye)

KABUL VE ONAY SAYFASI

Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Yapay Zeka ve Robotik Anabilim Dalı'nda Dr. Öğretim Üyesi Erol KINA danışmanlığında, Hayrullah TEMEL tarafından sunulan “Duygu Analizinde Transformer Tabanlı Modellerin Karşılaştırılması” başlıklı bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 19 /06 /2025 tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve yüksek lisans tezi olarak kabul edilmiştir.

Başkan: Dr. Öğretim Üyesi İsa AVCI

İmza:

Üye: Dr. Öğretim Üyesi Erol KINA

İmza:

Üye: Dr. Öğretim Üyesi Murat KOCA

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun /.... /..... tarih ve sayılı kararı ile onaylanmıştır.

İmza

Enstitü Müdürü

ETİK BEYAN

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

İmza

Hayrullah TEMEL



ÖZET

DUYGU ANALİZİNDE TRANSFORMER TABANLI MODELLERİN KARŞILAŞTIRILMASI

TEMEL, Hayrullah

Yüksek Lisans Tezi, Yapay Zeka ve Robotik Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Erol KINA

Temmuz 2025, 63 sayfa

Duygu analizi metinlerden duygu çıkarımı yapan yapay zeka çalışmasıdır. Son dönemlerde bu alanda yapılan çalışmalar ve alana olan rağbet bu alan ile ilgili çalışmaları da artırmıştır. Çeşitli makine öğrenmesi algoritmaları, derin öğrenme yöntemleri ve son zamanlarda Transformer tabanlı modeller duygu analizi konularında oldukça yaygın bir şekilde kullanılmaya başlamıştır.

Bu çalışma duygu analizi problemlerinde kullanılmak üzere Transformer tabanlı modellerin karşılaştırılması için yapılmıştır. Çalışmada biri 61121 satır ve diğeri 5937 satırdan oluşan iki ayrı veri seti kullanılmıştır.

Araştırmada kullanılan veri setleri öncelikli olarak temizlik ve ön hazırlık aşamalarından geçirilmiş ve Transformer tabanlı modellerin kullanımı için hazır hale getirilmiştir. Daha sonra on ayrı Transformer modeli için ayrı ayrı notebooklar çalıştırılmış ve sonuçları raporlanmıştır. Her bir notebook beş epoch boyunca çalıştırılmış ve hem uzun veri seti için hem de az örneklem sayısına sahip veri seti için sonuçlar karşılaştırılmıştır.

Çalışma sonucunda Transformer tabanlı modellerin performansı karşılaştırılmış olup çeşitli parametre ve metrik kullanımları durumundaki sonuçları gözlemlenmiştir. Genel olarak yüksek orandaki başarı performansları ve hızlı çalışabilme yetenekleri Transformer tabanlı modellerin duygu analizi çalışmalarında büyük oranda kullanılması gerektiğini göstermiştir.

Böylece diğer birçok Doğal Dil İşleme (NLP) görevinde olduğu gibi duygu analizi çalışmalarında da Transformer tabanlı modeller başarılı sonuçlar vermeye devam edecektir. Daha fazla parametre ayarı ve modele özgü ince ayarlarla gelecekte yüksek performanslı modeller kullanılabilir olacak olup gerçek dünya problemlerinde de çok daha büyük başarılarla imza atılabilecektir.

Anahtar kelimeler: Doğal dil işleme, Duygu analizi, Transformer modeller



ABSTRACT

COMPARISON OF TRANSFORMER-BASED MODELS IN SENTIMENT ANALYSIS

TEMEL, Hayrullah

M.Sc. Thesis, Department of Artificial Intelligence and Robotics

Supervisor: Asst. Prof. Dr. Erol KINA

July 2025, 63 pages

Sentiment analysis is an artificial intelligence task that involves extracting emotions from texts. In recent years, the increasing interest and studies in this field have significantly contributed to the development of sentiment analysis research. Various machine learning algorithms, deep learning methods, and more recently, transformer-based models have become widely used in sentiment analysis tasks.

This study was conducted to compare transformer-based models for use in sentiment analysis problems. Two separate datasets were used in the study: one consisting of 61121 entries and the other consisting of 5937 entries.

The datasets used in the research were first subjected to cleaning and preprocessing steps to prepare them for use with transformer-based models. Subsequently, separate notebooks were run for ten different transformer models, and their results were documented. Each notebook was run for five epochs, and results were compared for both the large dataset and the smaller dataset with fewer samples.

As a result of the study, the performance of transformer-based models was compared, and their outcomes under various parameters and metric usages were observed. In general, the high performance and fast processing capabilities of transformer models have demonstrated that they should be widely employed in sentiment analysis tasks.

Thus, just as in many other Natural Language Processing (NLP) tasks, transformer-based models are expected to continue delivering successful results in sentiment analysis as well. With further parameter tuning and model-specific optimizations, highly performant models could be utilized in the future, achieving even greater success in real-world applications.

Keywords: Natural language processing, Sentiment analysis, Transformer models



TEŞEKKÜR

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen danışmanım Sayın Dr. Öğretim Üyesi Erol KINA'ya teşekkür ederim. Ayrıca maddi ve manevi desteklerini esirgemeyen aileme teşekkürlerimi sunarım.

2025

Hayrullah TEMEL





İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vii
ÇİZELGELER LİSTESİ.....	ix
ŞEKİLLER LİSTESİ.....	xi
SİMGELER VE KISALTMALAR	xiii
1. GİRİŞ.....	1
2. LİTERATÜR BİLDİRİŞLERİ	5
2.1 Çalışmada Kullanılan Transformer Modelleri ve Çeşitli Kavramlar	10
2.1.1 NLP (Natural Language Processing-Doğal Dil İşleme)	10
2.1.2 Duygu Analizi	12
2.1.3 Transformer Mimarisi	13
2.1.3.1 BERT (Bidirectional Encoder Representations from Transformers)	15
2.1.3.2 RoBERTa (Robustly Optimized BERT Approach).....	15
2.1.3.3 DistilBERT (Distilled BERT).....	16
2.1.3.4 ALBERT (A Lite BERT).....	17
2.1.3.5 DeBERTa (Decoding-enhanced BERT with Disentangled Attention)	17
2.1.3.6 XLNet	18
2.1.3.7 GPT (Generative Pre-trained Transformer).....	19
2.1.3.8 ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)	19
2.1.3.9 T5 (Text-to-Text Transfer Transformer)	20
2.1.3.10 BART (Bidirectional and Auto-Regressive Transformer)	20
2.2 Accuracy (Doğruluk).....	21
2.3 Classification Report	21
2.4 Precision (Kesinlik).....	22
2.5 Recall (Duyarlılık).....	22
2.6 F1-Score	23
2.7 Support	23
2.8 Confusion Matrix (Karmaşıklık Matrisi)	23
3. MATERYAL VE YÖNTEM.....	25
3.1 Veri Seti.....	25

3.1.1	Twitter (X) Veri Seti	25
3.1.2	Emotion Veri Seti.....	28
3.2	Python.....	30
3.3	Google Colaboratory (Colab).....	31
3.4	Tokenizasyon.....	31
3.5	Transformer Tabanlı Modellerin Eğitimi ve Çalıştırılma Yöntemi	32
4.	BULGULAR.....	35
4.1	Kullanılan Modeller	35
4.1.1	BERT	35
4.1.2	DistilBERT.....	37
4.1.3	DeBERTa.....	37
4.1.4	RoBERTa.....	39
4.1.5	ALBERT	40
4.1.6	ELECTRA.....	42
4.1.7	GPT	43
4.1.8	T5	44
4.1.9	BART	45
4.1.10	XLNet.....	47
4.2	Değerlendirme	48
4.2.1	Twitter (x) Veri Seti.....	48
4.2.2	Emotion Veri Seti.....	51
5.	TARTIŞMA VE SONUÇ	55
	KAYNAKLAR.....	59
	ÖZ GEÇMİŞ.....	63

ÇİZELGELER LİSTESİ

	Sayfa
Çizelge 2.1 Classification Report örneği	22
Çizelge 3.1 Twitter (x) veri setindeki ilk beş satır gösterimi	26
Çizelge 3.2 Emotion veri setindeki ilk beş satır gösterimi	29
Çizelge 4.1 BERT modeli için Twitter (x) veri seti ile alınan sonuçlar	36
Çizelge 4.2 BERT modeli için Emotion veri seti ile alınan sonuçlar	36
Çizelge 4.3 DistilBERT modeli için Twitter (x) veri seti ile alınan sonuçlar	37
Çizelge 4.4 DistilBERT modeli için Emotion veri seti ile alınan sonuçlar	37
Çizelge 4.5 DeBERTa modeli için Twitter (x) veri seti ile alınan sonuçlar	38
Çizelge 4.6 DeBERTa modeli için Emotion veri seti ile alınan sonuçlar	38
Çizelge 4.7 RoBERTa modeli için Twitter (x) veri seti ile alınan sonuçlar	39
Çizelge 4.8 RoBERTa modeli için Emotion veri seti ile alınan sonuçlar	40
Çizelge 4.9 ALBERT modeli için Twitter (x) veri seti ile alınan sonuçlar	41
Çizelge 4.10 ALBERT modeli için Emotion veri seti ile alınan sonuçlar	41
Çizelge 4.11 ELECTRA modeli için Twitter (x) veri seti ile alınan sonuçlar	42
Çizelge 4.12 ELECTRA modeli için Emotion veri seti ile alınan sonuçlar	42
Çizelge 4.13 GPT modeli için Twitter (x) veri seti ile alınan sonuçlar	43
Çizelge 4.14 GPT modeli için Emotion veri seti ile alınan sonuçlar	44
Çizelge 4.15 T5 modeli için Twitter (x) veri seti ile alınan sonuçlar	44
Çizelge 4.16 T5 modeli için Emotion veri seti ile alınan sonuçlar	45
Çizelge 4.17 BART modeli için Twitter (x) veri seti ile alınan sonuçlar	46
Çizelge 4.18 BART modeli için Emotion veri seti ile alınan sonuçlar	46
Çizelge 4.19 XLNet modeli için Twitter (x) veri seti ile alınan sonuçlar	47
Çizelge 4.20 XLNet modeli için Emotion veri seti ile alınan sonuçlar	47
Çizelge 4.21 Twitter (x) veri setinde Transformer modellerinin negatif, pozitif ve nötr sınıflarında precision, recall ve f1-score oranı karşılaştırması	51
Çizelge 4.22 Emotion veri setinde Transformer modellerinin anger, joy ve fear sınıflarında precision, recall ve f1-score oranı karşılaştırması	54

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 1.1 NLP görevleri	1
Şekil 1.2 Duygu sınıflandırmasının emojilerle gösterimi.....	2
Şekil 2.1 NLP çalışma düzeni	11
Şekil 2.2 Duygu analizi	12
Şekil 2.3 Transformer mimarisinin çalışma mantığı	14
Şekil 2.4 BERT modelinin giriş gösterimi	15
Şekil 2.5 RoBERTa mimarisinin çalışma mantığı	16
Şekil 2.6 XLNet mimarisinin çalışma mantığı	18
Şekil 2.7 Confusion Matrix örneği	23
Şekil 3.1 Twitter (X) veri setinde temizlik adımları uygulanmadan önceki kelime bulutu	26
Şekil 3.2 Twitter (X) veri setinde temizlik adımları uygulandıktan sonraki kelime bulutu	27
Şekil 3.3 Emotion veri setindeki ilk beş satır gösterimi	28
Şekil 3.4 Emotion veri setindeki Emotion sütununun etiket dağılımı	29
Şekil 3.5 Emotion veri setinde temizlik adımları uygulanmadan önceki kelime bulutu	30
Şekil 3.6 Emotion veri setinde temizlik adımları uygulandıktan sonraki kelime bulutu	30
Şekil 4.1 BERT modeli için karmaşıklık matrisleri	36
Şekil 4.2 DeBERTa modeli için karmaşıklık matrisleri	39
Şekil 4.3 RoBERTa modeli için karmaşıklık matrisleri	40
Şekil 4.4 ALBERT modeli için karmaşıklık matrisleri	41
Şekil 4.5 ELECTRA modeli için eğitim ve validasyon kayıp grafikleri	43
Şekil 4.6 T5 modeli için epoch başına ortalama kayıp grafikleri	45
Şekil 4.7 BART modeli için karmaşıklık matrisleri	46
Şekil 4.8 XLNet modeli için karmaşıklık matrisleri	48
Şekil 4.9 Twitter (x) veri setinde negatif sınıfının değerlendirme metrikleri oranları ..	49
Şekil 4.10 Twitter (x) veri setinde pozitif sınıfının değerlendirme metrikleri oranları .	50
Şekil 4.11 Twitter (x) veri setinde nötr sınıfının değerlendirme metrikleri oranları	50

Şekil 4.12 Emotion veri setinde anger sınıfının değerlendirme metrikleri oranları	52
Şekil 4.13 Emotion veri setinde joy sınıfının değerlendirme metrikleri oranları	52
Şekil 4.14 Emotion veri setinde fear sınıfının değerlendirme metrikleri oranları	53



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamalarıyla aşağıda sunulmuştur.

Kısaltmalar	Açıklama
ALBERT	A Lite BERT
BART	Bidirectional and Auto-Regressive Transformer
BERT	Bidirectional Encoder Representations from Transformers
Colab	Colaboratory
DeBERTa	Decoding-enhanced BERT with Disentangled Attention
DistilBERT	Distilled BERT
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
FN	False Negative
FP	False Positive
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
MLM	Maskeli Dil Modelleme
NER	Adlandırılmış Varlık Tanıma
NLP	Doğal Dil İşleme
RoBERTa	Robustly optimized BERT approach
T5	Text-To-Text Transfer Transformer
TN	True Negative
TP	True Positive
TPU	Tensor Processing Unit



1. GİRİŞ

Günümüzde teknolojinin büyük bir hızla gelişmesi ve buna bağlı olarak internet kullanımının yaygınlaşması, dijital mecralarda her gün çok sayıda metin tabanlı içeriğin üretilmesine sebep olmuştur. Sosyal medya, müşteri yorumları ve çevrimiçi forumlar gibi platformlar, insanların fikirlerini ve duygularını ifade etmeleri için önemli bir mecra hâline gelmiştir. Bu büyük orandaki verilerin analiz edilmesi, insanların duygu durum ifadelerini anlamak ve sınıflandırmak için yeni yaklaşımların geliştirilmesini zorunlu kılmıştır. Burada duygu analizi çalışmaları, doğal dil işleme (Natural Language Processing- NLP) alanında önemli bir alan olmuştur. Bu doğrultuda, duygu analizi çalışmaları NLP'nin alt alanlarından biri olarak metinlerdeki anlamı daha derinlemesine çözümlemek için sıklıkla kullanılmaktadır (Şekil 1.1).



Şekil 1.1 NLP görevleri (Anonim, 2022)

Duygu analizi, metinlerden veya metin tabanlı içeriklerden insanların duygu durumlarını (örneğin negatif, nötr ve pozitif) saptamayı öncelikleyen bir sınıflandırma çalışması alanıdır. Bu alan müşteri bildirimleri, toplumdaki eğilimlerin tespit edilmesi, model analizi yapılması vb. birçok uygulama alanını içerir. Fakat metin tabanlı duyguların analizi dil yapısının karmaşıklığı, birden fazla anlamlılık, bağlama olan bağımlılığı ve verinin çeşitliliği gibi zorlukları da barındırmaktadır.

Diğer bir ismi de Fikir Madenciliği (Opinion Mining) olan duygu analizi, belirli konu hakkında bir cümlelerin ya da textin duygu içerip içermediğini anlamak için yapılan çalışmalar bütünüdür (Barak, 2021). Doğal Dil İşleme alanında sık kullanılan bir yöntem olan duygu sınıflandırması çalışmaları resim ve metinlerden duygu çıkarımı şeklinde de yaygınlık kazanmakta olup bu metin veya görseller birçok farklı konuda olabilmektedir. Duygu analizinde duygu içeren ifadeler makine diline çevrildikten sonra işlenir ve çeşitli duyguları açıklayan görsellerle veya ifadelerle sunulur. Şekil 1.2’de bu görsellere bir örnek verilmiştir.



Şekil 1.2 Duygu sınıflandırmasının emojilerle gösterimi (Özdamar, 2020)

Medya sektöründen siyaset arenasına kadar birbirinden ayrı birçok alanda kullanılan duygu sınıflandırması sayesinde bireylerin veya toplumların çeşitli konular hakkında fikirleri olabilmekte ve buna göre kararlar alınabilmektedir. Metinlerdeki veya görsellerdeki duygu analizinin başarılı sonuçlar vermesi bu sonuçları kullanan insanlar açısından da verimli olmaktadır. Bu sayede istenilen yönlendirmeler yapılabilmekte ve insanların buna göre pozisyon almaları sağlanabilmektedir. Bu durumda duygu

analizinin doğruluk değeri de önem kazanmaktadır. Yüksek başarı oranlarındaki duygu analizi çalışmaları gerçek dünya verilerine uygulanabilirliği artırma olanağına sahiptir. Bu doğrultuda, duygu analizinin doğruluk oranını artırmak ve daha karmaşık dil yapılarını anlayabilmek amacıyla, geleneksel yöntemlerin ötesine geçen yeni yaklaşımlar geliştirilmektedir. Özellikle derin öğrenme temelli modeller, doğal dil işleme alanında önemli ilerlemeler sağlamış ve duygu sınıflandırmasında daha başarılı sonuçlar elde edilmesine olanak tanımıştır.

Son dönemde klasik makine öğrenmesi yöntemlerine göre daha avantajlı bir konuma gelen derin öğrenme tabanlı modellemeler duygu analizinde önemli yenilikler ve gelişmeler sunmuştur. Özellikle Transformer tabanlı model çalışmaları bu alanda devrim niteliğinde çalışmalar kazandırmıştır. Transformer mimarisinin çift yönlü bağlam analizi yapabilmesi, aynı anda paralel işlem yapabilme kapasitesi ve büyük veri setleri üzerinde güçlü temsil oranlarına sahip olması, doğal dil işleme problemlerinde klasik makine öğrenmesi yöntemlerine göre üstün başarılar elde etmesini sağlamıştır. BERT, DistilBERT, GPT, ELECTRA, RoBERTa, XLNet vb. gibi modeller metni anlama ve sınıflandırma çalışmalarında yüksek başarı oranlarını yakalayarak duygu analizinde çokça kullanılmaktadır. Çeşitli çalışmalarda hibrit yaklaşımlar da ele alınmakta ve klasik makine öğrenmesi yöntemleriyle gerek derin öğrenme yaklaşımları gerekse de Transformer tabanlı modellerle karşılaştırılmaları yapılmakta ve başarı oranları yükseltilmektedir. Tüm bu gelişmeler Transformer tabanlı modellerin duygu analizinde sunduğu avantajların sistematik bir şekilde değerlendirilmesini gerekli kılmaktadır. Bu kapsamda farklı Transformer mimarilerinin çeşitli metrikler doğrultusunda karşılaştırılması, hangi modelin hangi veri setlerinde daha etkili sonuçlar verdiğini ortaya koymak açısından önem taşımaktadır. Bu bağlamda sunulan tez çalışmasında Transformer tabanlı modellerin duygu analizi alanındaki başarılarının karşılaştırılması amaçlanmıştır. Bu doğrultuda farklı Transformer modelleri belirli veri setleri üzerinde eğitilmiş ve doğruluk, kesinlik, duyarlılık gibi performans metrikleriyle değerlendirilmiştir. Bu amaç doğrultusunda çalışmada izlenen yöntemler, kullanılan veri setleri ve model mimarilerine ilişkin teknik ayrıntılar ilerleyen bölümlerde sunulmuştur.

Bu tez Transformer modellerinin duygu analizi çalışmalarındaki başarılarını karşılaştırmayı hedeflemektedir. Çalışmada farklı Transformer modelleri belirli veri

setleri üzerinde eğitilmiş ve doğruluk, kesinlik, duyarlılık gibi metriklerle değerlendirilmiştir. Varılan sonuçlar duygu analizi yapmak için hangi modellerin daha uygun olduğunu ortaya koymaktadır.

Çalışmanın ileriki bölümlerinde duygu analizi ve Transformer tabanlı modeller hakkında kapsamlı bilgiler sunulmakta, bu bağlamda kullanılan yöntemler açıklanmakta ve elde edilen sonuçlar değerlendirilmektedir.



2. LİTERATÜR BİLDİRİŞLERİ

Bu bölümde tez çalışması ile alakalı yapılan literatür taraması çalışmaları yer almaktadır. Yapılan literatür taraması şu şekilde özetlenebilir.

Almufareh vd. (2024) tarafından yapılan çalışmada tweet verileri üzerinde penta-class sınıflandırması gerçekleştirilmiş ve bu amaçla Bert tabanlı BertSent modeli kullanılmıştır. Çalışma sonucunda model, sırasıyla %78.3 ve %75.3 doğruluk oranları elde etmiştir. Elde edilen bulgular, BertSent modelinin çift yönlü öğrenme tekniği kullanan diğer modellere kıyasla daha başarılı olduğunu ortaya koymuştur (Almufareh vd., 2024).

BERT ve RoBERTa modellerinin, özellikle gelişmiş sinir ağı mimarileriyle birlikte kullanıldıklarında iyi sonuçlar verdiği, RoBERTa modelinin çeşitli yapılandırmalarda rekabetçi sonuçlar ürettiği belirtilmiştir (Jahin vd., 2024).

BERT, ALBERT, RoBERTa, DistilBERT ve XLNet gibi çeşitli Transformer tabanlı modeller üzerinde yapılan denemeler sonucunda, %89.16 doğruluk oranıyla RoBERTa modelinin yön tabanlı duygu analizinin karmaşıklığını azaltmada en başarılı sonucu verdiği bulunmuştur. Çalışma, Transformer tabanlı modellerin yön tabanlı duygu analizi konusunda başarılı sonuçlar ürettiğini gösterse de bazı sınırlamalarının da olduğu gösterilmiştir. Bu modellerin daha fazla ince ayar yapılarak kullanılmaları durumunda ise etkinliklerinin artabileceği ifade edilmiştir (Perikos ve Diamantopoulos, 2024).

Aljabar vd. (2024) BERT ile 5003 satırlık veri seti üzerinde yaptıkları eğitimde %98 gibi bir doğruluk elde ederek modelin güçlülüğünü ortaya koymuşlardır. Modelin hata oranlarını azaltma ve eğitim sürecinde hızlı yakınsama kapasitesi, yalnızca %3'lük düşük bir eğitim kaybıyla gösterilmiştir.

Fan vd. (2021) tarafından yapılan çalışmada BERT, DistilBERT ve RoBERTa modelleri karşılaştırılmış ve çalışma sonucunda BERT modelinin diğer iki modele kıyasla en iyi sonuçları verdiği bulunmuştur. Çalışmada BERT modelinin toksik yorumları yüksek seviyede doğruluk oranlarıyla bulduğu ve sınıflandırma yapabildiği gözlemlenmiştir. Çalışmada kullanılan tweetlerin elle etiketlenmesi ve daha fazla sayıda örneklem olması durumunda emojilerin de hesaba katılabileceği çalışma sonunda vurgulanmıştır.

BERT ve ELECTRA modelleri üzerinde yapılan bir çalışmada tam eğitim aldıkları taktirde ELECTRA modelinin BERT'ten daha yüksek aşağı akış doğruluğu elde ettiği belirtilmiştir. Ayrıca ELECTRA modelinden elde edilen sonuçların modeller küçüldükçe büyüdüğü ifade edilmiştir (Clark vd., 2020).

Ethan (2025) çalışmasında duygu analizinde Bi-LSTM, BERT ve RoBERTa gibi modellerin kapsamlı bir değerlendirmesini gerçekleştirmiştir. Çalışma bulgularına göre, Transformer tabanlı modellerin Bi-LSTM'ye kıyasla genel olarak daha avantajlı olduğu, ancak kaynakların kısıtlı olduğu ortamlarda dezavantaj oluşturabileceği vurgulanmıştır. Ayrıca, BERT ve RoBERTa modellerinin özellikle belirsiz ifadeler içeren cümlelerde veya spesifik kelimeler barındıran metinlerde daha düşük performans gösterebileceği ifade edilmiştir.

El Azzouzy vd. (2025) tarafından yürütülen çalışmada Youtube yorumları üzerinden BERT, GPT, RoBERTa ve T5 gibi temel Transformer modeller karşılaştırılmıştır. Çalışma bu modellerin güçlü yönlerini ve genel olarak duygu sınıflandırmasındaki başarı performanslarını ortaya koymuştur. Değerlendirilen bu modeller arasında RoBERTa modeli üstün bir performans göstererek yüksek doğruluk ve F1 puanları elde etmiştir. Çalışmada ayrıca Youtube gibi platformlardan alınan yorumların incelenmesinde spesifik ayarların yapılması durumunda çok daha başarılı sonuçlar elde edilebileceği vurgulanmıştır. Yazarlar bu çalışma sonunda aşırı uyum riski, değerlendirilen modellerin sınırlı çerçevesi ve bazı metriklerle bağlılık gibi sebepler yüzünden bulguların gerçek dünya verilerine uygulanabilirliğinin zorlaştığını ve bunun da GPT-4 veya hibrit modeller ile aşılabileceğini vurgulamışlardır (El Azzouzy vd., 2025).

Fadhil (2024) tarafından yapılan çalışmada Transformer modellerinden BERT ve DistilBERT modelleri karşılaştırılmıştır. Yazar, BERT modelinin DistilBERT gibi nispeten daha küçük modellere göre daha fazla eğitim süresine ihtiyaç duyduğunu belirtmiş; ancak elde edilen yüksek eğitim doğruluğunun bu dezavantajı ortadan kaldırdığını vurgulamıştır. DistilBERT modelinin ise büyük veri setleri veya sınırlı kaynaklar ile çalışmada avantaj sağladığını ve bu özellikleriyle öne çıktığını belirtmiştir. Yapılan çalışmada BERT modelinin 54 dakikada %91.39 gibi bir doğruluk değerine ulaştığını buna karşılık DistilBERT'in 39 dakika 15 saniyede %91.80 doğruluk oranına ulaştığını açıklamıştır. Genel olarak bu iki model arasındaki seçimin yüksek doğruluk

ihtiyacı ile daha hızlı ve düşük kaynak kullanımı olduğunu eklemiştir. Gelecek çalışmalarda bu modellerde daha fazla ince ayar yapılabileceğini ve erken durdurma, öğrenme hız ayarı gibi parametrelerin daha esnek kullanılabileceğini ifade etmiştir (Fadhil, 2024).

Singh (2024) tarafından yürütülen çalışmada, Transformer tabanlı modellerden BERT, GPT ve T5 üzerine karşılaştırmalı analizler gerçekleştirilmiştir. Çalışma sonucunda, BERT modeli %90 doğruluk (accuracy) oranıyla en yüksek performansı göstermiştir; onu sırasıyla %87 ile T5 ve %85 ile GPT modelleri takip etmiştir. Eğitim süresi açısından ise GPT modelinin diğer modellere göre daha kısa sürede eğitildiği gözlemlenmiştir. Bellek kullanımı bakımından da GPT, en az kaynak tüketen model olarak öne çıkmış; ardından BERT ve T5 modelleri gelmiştir.

Eser ve Sahin (2024) yaptıkları çalışmada Transformer modellerinden BERT, DistilBERT, RoBERTa, XLM-RoBERTa modellerini kullanıcı puanlarını tahmin etmede kullanmış ve karşılaştırma yapmışlardır. Çalışma sonunda DistilBERT modelinin hatırlama ve doğruluk bakımından diğer modellerden daha başarılı olduğu görülmüştür. En iyi F1 skorunu XLM-RoBERTa modeli almış ve kesinlik ve hatırlama gibi metriklerin dengelenmesinde başarılı sonuçlar vererek tutarlı bir sınıflandırma performansı sergilemiştir. XLM-RoBERTa modeli özellikle 5 sınıflarını yakalamada daha başarılı olup 1 ve 2 sınıflarını yakalamada kısmen zayıf kalmıştır.

Shetty vd. (2025) yaptıkları çalışmada DistilBERT, ALBERT, RoBERTa, BERT, ELECTRA ve XLNet gibi Transformer tabanlı modelleri hem kendi aralarında hem de hibrit modellerle karşılaştırmışlar ve detaylı sonuç raporları vermişlerdir. Çalışma sonunda XLNet mimarisinin amazon verilerinin iki ayrı kategorisinde %95 ve %95.2 gibi yüksek doğruluk oranlarıyla sonuçlar verdiğini ancak göz önünde bulundurulması gereken bazı dezavantajlı durumların da olduğunu vurgulamışlardır. Yüksek hesaplamalarda karmaşıklığın büyük miktardaki verilerde büyük hacimlerde işlemler gerektirebileceğini söylemişlerdir. XLNet mimarisinin hesaplama maliyetinin DistilBERT ve ELECTRA gibi modellere kıyasla fazla olduğunu ve bunun tolere edilemeyebileceğini belirtmişlerdir. Çalışmada ayrıca DistilBERT modelinin de kısa sürede eğitimi tamamladığını ve kaynakların kısıtlı olduğu durumlarda gayet kullanışlı olabileceğini de eklemiştir. Gelecek çalışmalarda çok dilli yorumların yönetilmesi ve

modellerin kısa ve belirsiz yorumları işleyebilme becerilerinin artırılmasının uygulanabilirliklerini de artırabileceğinin üzerinde durmuşlardır.

Yaptıkları çalışmada Hamed ve Zaidkilani, (2025) Nakba kısa hikayelerinin anlamsal benzerliğini ve yazar tarzlarını Transformer tabanlı BERT, RoBERTa ve AraBERT gibi modelleri kullanarak karşılaştırmışlardır. Yazarlar, bütün modellerin yüksek benzerlik oranları elde ettiğini ve bunun da modellerin incelikli dil yapısına sahip olmaları, kültürel bağlamlar ve Arapça dilindeki çeşitlilikleri yakalamadaki becerilerinden kaynaklandığını savunmuşlardır. RoBERTa modelinin BERT modelini biraz geride bırakarak daha üstün bir performans sergilediğini belirten araştırmacılar, kullandıkları Transformer modellerinin TF IDF kullanan geleneksel makine öğrenmesi algoritmalarını da büyük oranda geçtiğini ve bu modellerin dilin bağlamını anlamada üstün yeteneklere sahip olduklarını vurgulamışlardır (Hamed ve Zaidkilani, 2025).

Apu vd. (2025), Human Metapneumovirus (HMPV) virüsü ile ilgili çeşitli YouTube yorumlarını içeren bir veri seti üzerinden Transformer mimarilerini karşılaştırmışlardır. Çalışma sonucunda, XLNet modelinin BERT'e kıyasla uzun dizileri işleme kapasitesi ve çift yönlü bağlamı yakalamadaki başarısı açısından daha üstün olduğu belirtilmiştir. Permütasyon tabanlı eğitim yaklaşımından yararlanan XLNet'in, tokenlar arasındaki ilişkileri daha etkili şekilde modelleyebildiği ve bu özelliği sayesinde YouTube yorumları gibi kullanıcı tarafından üretilen gayri resmi dili analiz etmek için güçlü bir aday olduğu ifade edilmiştir. Yazarlar, XLNet'in bu performans düzeyine ince ayar (fine-tuning) ve ön eğitim süreçlerinin katkısıyla ulaştığını vurgulamıştır. Ayrıca çalışmada, kaynak kullanımıyla ilgili sınırlamalara da dikkat çekilmiş; yorumların karmaşık duygular içerebilmesinin bazı analizlerde zorluk yaratabileceği ifade edilmiştir. Gelecek çalışmalarda yalnızca metin değil, aynı zamanda video veya görsel içeriklerin de analiz sürecine entegre edilmesinin gerekliliği savunulmuştur

Riaz vd. (2025) tarafından yapılan çalışmada farklı Transformer mimarilerinin çeşitli yönleri karşılaştırmalı olarak analiz edilmiştir. Çalışma sonunda, BERT ve XLNet-BERT hibrit modelinin konu tutarlılığı bakımından öne geçtiği, konu çeşitliliği bakımından ise GPT-LDA'nın en yüksek puanla öne çıktığı ve XLNet-BERT'in onu takip ettiği belirtilmiştir. Çalışmanın sonunda genel olarak XLNet-BERT modelinin en başarılı model olduğu kanısına varılmış, ayrıca anlamsal ilişkileri yakalaması

bakımından BERT modelinin öne çıktığı belirtilmiştir. Yazarlar, bu sonuçların umut verici olduğunu belirterek gelecek çalışmalar adına daha fazla inceleme yapılmasını tavsiye etmişlerdir (Riaz vd., 2025).

Yaptıkları çalışmada Zhang ve Shafiq (2024) BERT, XLNet, RoBERTa, GPT-2 ve ALBERT modellerini duygu analizi, soru cevaplama, metin oluşturma ve özetleme ve isim varlık tanıma gibi konularda karşılaştırmış ve sonuçlar almışlardır. Duygu analizinde en yüksek performansı RoBERTa modelinin verdiğini, soru yanıtlama görevlerinde yine RoBERTa modelinin öne çıktığını ve BERT modelinin diğerleri kadar iyi olmadığını, adlandırılmış varlık tanıma görevlerinde RoBERTa modelinin diğerlerini geride bıraktığını, metin özetleme görevlerinde en yüksek doğruluk değerine XLNet modelinin ulaştığını ve RoBERTa modelinin bu alanda geride kaldığını, konu modelleme görevlerinde GPT-2 ve RoBERTa modellerinin daha başarılı olduğunu ve XLNet modelinin bu konuda en az başarıyı yakaladığını ve metin üretme görevlerinde ise özellikle RoBERTa modelinin öne çıktığını ve ardından GPT-2 modelinin de iyi performans sergilediğini saptamışlardır.

Shaik Vadla vd. (2024) yaptıkları çalışmada Transformer modellerinden olan BERT ve T5 modellerini karşılaştırmışlardır. Çalışmada BERT için %92 ve T5 için %91 oranında doğruluk değerleri bulmuşlardır. Çalışmada yazarlar Transformer modellerinden özellikle BERT modelinin güçlü çıkarım yapma yeteneklerinin olduğunu vurgulamışlar ve bunun da gelecek çalışmalar için çok umut verici bir durum olduğunu belirtmişlerdir. Bu modelin ayrıca tasarım içgörüsü çıkarmak için kullanılabileceğinden bahsetmişler ve bunun büyük bir fırsat olabileceğini söylemişlerdir.

Bu çalışmada Bandaranayake ve Usoof (2025), Sinhala haber yorumlarının duygu analizi için BERT, DistilBERT, RoBERTa ve XLM RoBERTa modellerini karşılaştırmışlardır. Çalışma sonunda BERT ve DistilBERT modellerinin üç ve dört sınıflı duygusal analiz çalışmalarında rekabetçi makro-F1 ve doğruluk sonuçları elde ettiğini fakat BERT, DistilBERT ve RoBERTa modellerinin makro-F1 değerlerinin XLM-R modellerinden daha düşük olduğunu belirtmişlerdir. Çalışmanın BERT ve DistilBERT modellerinin Sinhaladaki ilk deneysel olarak kullanıldıkları çalışma olduğunu belirten yazarlar transformer modellerinin küçük veri kümelerinde bile ince ayar yapılarak dikkate değer performanslar sergilediklerini vurgulamışlardır. Çalışma

ayrıca özellikle sınırlı kaynaklara sahip diller için bu modellerin çok avantajlı olduklarını da vurgulamaktadır (Bandaranayake ve Usoof, 2025).

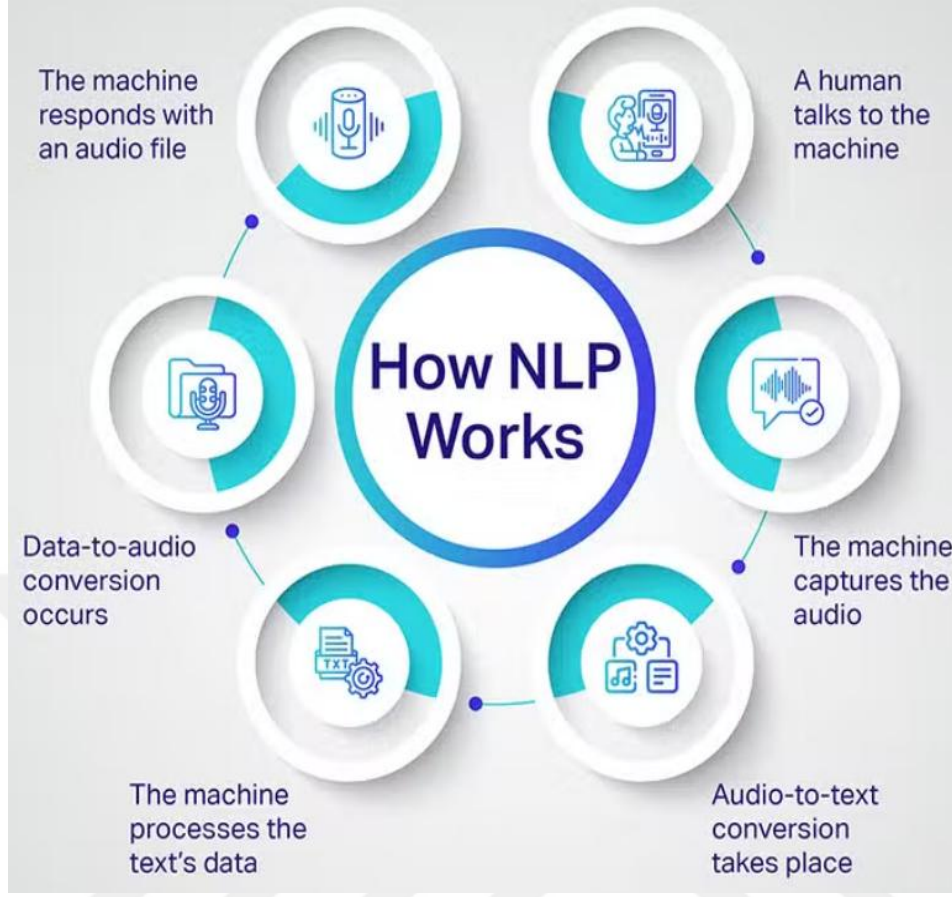
2024 yılında Aydın vd. (2024) tarafından yürütülen çalışmada, geleneksel makine öğrenmesi algoritmaları ile DistilBERT, BERT, GPT-2, RoBERTa ve GPT-3 gibi çeşitli Transformer mimarilerinin performansları karşılaştırılmıştır. Çalışma sonunda Transformer modelleri arasında GPT-3 modeli hem makine öğrenmesi modellerini hem de diğer Transformer tabanlı modelleri geçerek %77 oranında bir doğruluk ve F1-score performansı sergilemiştir. Çalışmanın yazarlarına göre Transformer mimarileri özellikle de GPT mimarisine dayananlar geniş veri setlerinde ve karmaşık metinlerde kapsamlı bir öğrenme performansı sergilemektedir ve bu da gelecek çalışmalar için umut verici olmaktadır.

2.1 Çalışmada Kullanılan Transformer Modelleri ve Çeşitli Kavramlar

Bu bölümde tezde kullanılan Transformer modelleri, çeşitli kavramlar, yapay zeka terimleri vb. için detaylı bilgiler verilmiştir.

2.1.1 NLP (Natural Language Processing-Doğal Dil İşleme)

NLP makinelere insan dilini öğretmeyi ve onları insan gibi düşündürmeyi amaçlayan bir disiplindir. Makine-insan etkileşimi için makine diline uygun veri hazırlamayı, bu verileri makinelerin anlayabileceği şekle sokmayı ve bunun sonucunda insana ait olan duyguları makine üzerinden okumayı konu edinir. Chatbot, sanal asistanlar, duygu analizi, metin sınıflandırma, özetleme, dil modelleri, makine çevirisi ve NER gibi yaygın bir yelpazesi vardır (Şekil 2.1).



Şekil 2.1 NLP çalışma düzeni (Anonim, 2025)

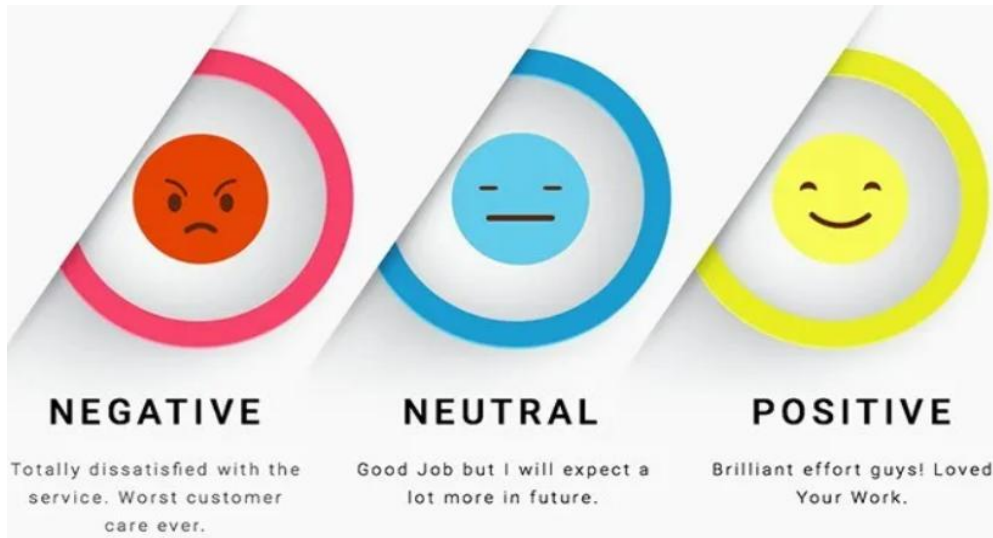
Doğal Dil İşleme insanların makineler arasındaki iletişimin gelişmesine katkı sunarak ve aralarında bilgi etkileşimi kurarak hem insanların hem de makinelerin performansını artırır. İş dünyasında kişisel asistanlar, sanal uygulamalarda chatbotlar, makine dilinin insan diline yakın bir seviyeye gelmesi NLP çalışmaları sayesinde gelişmiş ve büyük avantajlar sağlamıştır. Bu uygulamalar sayesinde teknolojinin günlük hayattaki etkisi artmış, insanlar ve makineler arasındaki iletişim kaynakları ve bilgi akışı çok daha verimli bir hale gelmiştir. NLP sayesinde dokümantasyon verimliliği artmış, özetleme ve çeviri gibi fonksiyonlar kısa zamanda içerisinde yapılabilir olmuş, metin madenciliği görevlerinde başarılı sonuçlar elde edilmiş ve insan verimliliği üst noktalara taşınmıştır (Anonim, 2025).

Dijital verinin çok fazla olması, internetin çok fazla yaygınlaşması ve iş süreçlerinin büyük bir kısmının bu mecraya taşınmış olması bu büyük verinin yorumlanması, düzenlenmesi gibi gereklilikleri de beraberinde getirmiştir. Bu yüzden NLP'ye daha fazla ihtiyaç duyulmaktadır. Yapay zekadaki son gelişmeler bunu artık

mümkün kılmaktadır. Kognitif Hesaplama, insanlarla doğal ve anlamlı bir şekilde öğrenebilen, akıl yürütebilen ve etkileşime girebilen sistemler oluşturmayı amaçlarken NLP makinelerin insan dilini anlamasını ve üretmesini sağlamaya odaklanır (Orynbay vd., 2024). Bilişsel hesaplama sistemlerinin geliştirilmesi insan zekası yerine geçebilecek sistemler geliştirilmesine katkı sunmuştur. Bu iki alanın teması, makinelerin insanlarla etkileşim kurma biçiminde devrim yaratma potansiyeline sahiptir ve insan dilini benzeri görülmemiş bir doğruluk ve bağlam farkındalığı ile anlayabilen, yorumlayabilen ve yanıt verebilen daha sezgisel ve akıllı sistemler sunar (Orynbay vd., 2024).

2.1.2 Duygu Analizi

Duygu analizi metinlerdeki duyguları, duygusal tonları analiz eden bir NLP alanıdır. Yerine göre metinleri pozitif, negatif ve nötr gibi duygu sınıflarıyla sınıflandırmak için (Şekil 2.2) bazen de mutluluk, sevinç, öfke, kızgınlık vb. daha geniş anlamda sınıflandırmalar için kullanılır. Müşteri geri bildirimleri, sosyal medya yorumları, sağlık, hukuk, finans, siyaset vb. alanlarda yaygın olarak kullanılır. Duygu analizi, pazarlamadan politikaya kadar çok sayıda uygulamaya sahiptir ve son on yılda giderek daha popüler bir araştırma konusu haline gelmiştir (Chifu ve Fournier, 2023)



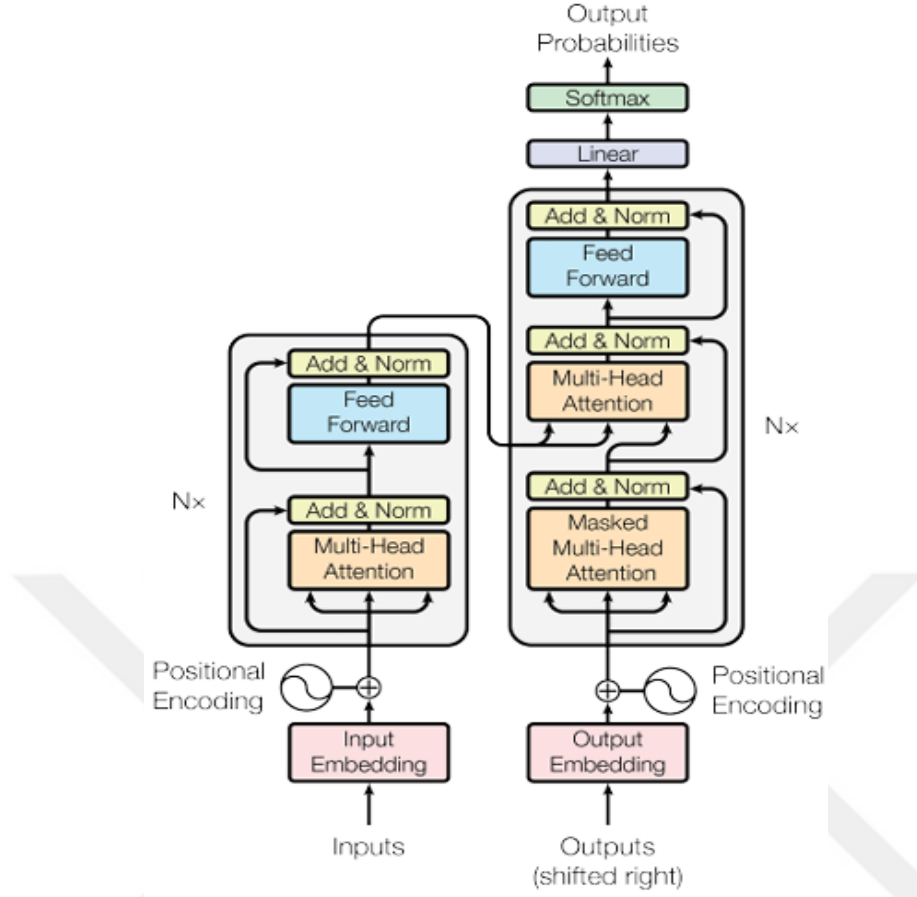
Şekil 2.2 Duygu analizi (Demir, 2019)

Duygu analizi metinlerden otomatik olarak bilgi çıkarma konusunda avantajlı bir araçtır. Duygu analizi, diğer uygulamaların yanı sıra müşterilerin belirli bir ürüne yönelik duygularını, seçmenlerin bir siyasi adaya yönelik görüşlerini veya hastaların tıbbi durumlarına yönelik duygularını belirlemek için kullanılabilir (Chifu ve Fournier, 2023).

Son yıllarda duygu analizi konusundaki başarılarla rağmen çeşitli sebeplerden ötürü duygu analizi yapma noktasında karşımıza çeşitli zorluklar da çıkmaktadır. Dildeki belirsizlik, eş anlamlılık veya farklı kültürlerde kelimelerin farklı anlamlara karşılık gelmesi gibi sebeplerle metinde amaçlanan anlam net olarak okunamayabilmektedir. Örneğin, "cool beans" ifadesi Amerikan İngilizcesinde "mükemmel" anlamına gelir, ancak ana dili İngilizce olmayanlar için anlamsız veya kafa karıştırıcı olabilir (Chifu ve Fournier, 2023)

2.1.3 Transformer Mimarisi

İlk kez 2017 yılında ‘Attention Is All You Need’ adlı makalede yayınlanan ve (Vaswani vd., 2017) tarafından önerilen Transformer mimarisi özellikle NLP çalışmalarında devrim niteliğinde bir adım olmuştur. Daha sonra yapılan çeşitli çalışmalarla bu modeller türetilmiş ve geliştirilmiştir. Son dönemlerde duygu sınıflandırmasından metin çözümlemelerine kadar bütün NLP çalışmalarında çok büyük bir oranda kullanılmaktadırlar. Özellikle çift yönlü bağlamı anlamada çok etkili olan bu model mimarileri encoder-decoder çalışma mantığıyla anlamsal bütünlemede oldukça etkili performans sergilemektedirler. Şekil 2.3’te Transformer modellerin çalışma mantığı verilmiştir.



Şekil 2.3 Transformer mimarisinin çalışma mantığı (Vaswani vd., 2017)

Transformer uzun vadeli ilişkileri kolayca ele alırken diziden diziye sorunları çözmeyi amaçlayan ayırt edici bir çerçevedir (Almufareh vd., 2024). Evrişim katmanı haricinde bilgi işleme süreçlerini dikkat mekanizması üzerine kurar. Kodlayıcı-kod çözücü metodolojisi kullanılarak oluşturulmuştur (Almufareh vd., 2024). Encoder-decoder dil bağlamı ile giriş bilgilerini analiz etmek için kodlayıcı (encoder) ve sonuç üretmek için de yorumlayıcı (decoder) kullanır.

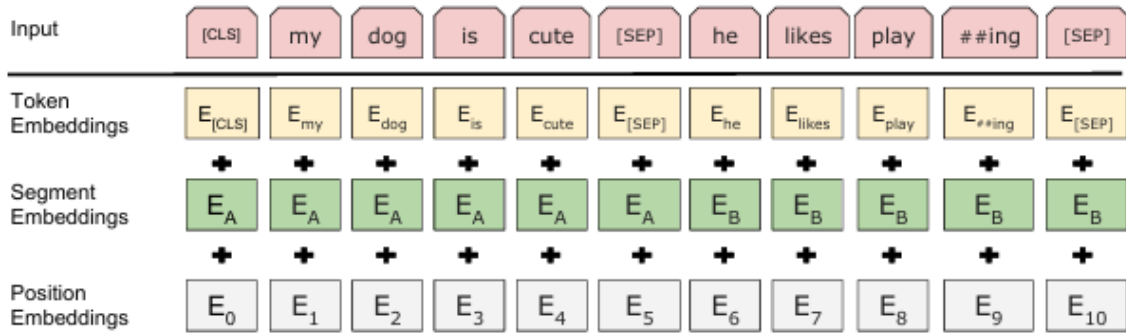
Transformer, önemli ölçüde daha fazla paralel işlemeye olanak tanır ve sadece sekiz P100 GPU üzerinde on iki saat kadar kısa bir süre eğitildikten sonra çeviri kalitesinde yeni bir en iyi duruma ulaşabilir (Vaswani vd., 2017). Transformer modelleri özellikle metinlerde anlam bütünlüğünün bozulmasına yol açan anlam kayıpları konusunda encoder-decoder özelliklerinin yanısıra kullandıkları attention (dikkat) mekanizmasıyla da bu konuda etkili sonuçlar sağlamaktadırlar. Bir dikkat fonksiyonu, bir sorguyu ve bir dizi anahtar-değer çiftini bir çıktıya eşleme olarak

tanımlanabilir; burada sorgu, anahtarlar, değerler ve çıktı hepsi vektörlerdir (Vaswani vd., 2017).

2.1.3.1 BERT (Bidirectional Encoder Representations from Transformers)

BERT 2018 yılında Google tarafından geliştirilen Transformer tabanlı bir dil modelidir. Doğal Dil İşleme (NLP) alanında yüksek başarı sağlayan bu dil modeli duygu sınıflandırması çalışmalarında da yüksek verimlilik sağlamıştır. BERT modeli, cümleleri hem sağdan sola hem de soldan sağa değerlendirmektedir. Böylece metindeki anlamı ve kelimelerin birbirleriyle olan ilişkilerini daha iyi çıkarmayı planlamakta ve sonuçlarda da buna göre yüksek başarı oranları elde edilmektedir (Devlin vd., 2018).

BERT birçok göreve özgü mimari oluşturan, cümle düzeyinde ve belirteç düzeyinde görevlerden oluşan geniş bir pakette son teknoloji performans elde eden ilk ince ayar tabanlı temsil modelidir (Devlin vd., 2018). Bu tanımdan da anlaşılacağı üzere BERT mimarisi birçok çalışmada başarılar elde edebilen ve özellikle de duygu analizi konusunda bağlam kapsayıcılığı üzerinde büyük başarı oranlarına ulaşabilen bir Transformer modelidir (Şekil 2.4).



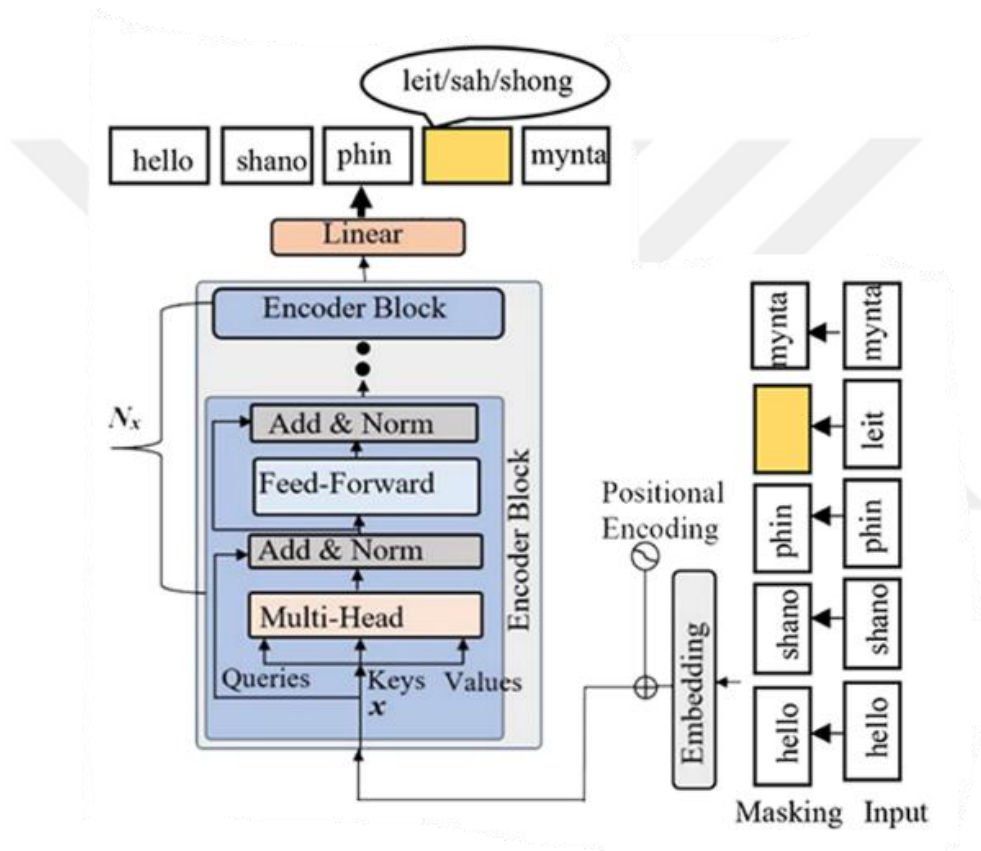
Şekil 2.4 BERT modelinin giriş gösterimi (Devlin vd., 2018b)

2.1.3.2 RoBERTa (Robustly Optimized BERT Approach)

Meta AI tarafından geliştirilen RoBERTa BERT modelinin bazı eksiklerini gidermek ve başarısını arttırmak için geliştirilmiş ileri bir versiyonudur. BERT'e bazı değişikliklerle dayanır, bunlar şunları içerir: bir sonraki cümle tahmin hedefini ortadan

kaldırma, daha uzun diziler üzerinde eğitim, eğitim verilerinde kullanılan maskeleme desenini dinamik olarak değiştirme ve daha büyük gruplar üzerinde ve daha fazla veri arasında daha uzun eğitim (Mitri vd., 2024).

RoBERTa mimarisi ön eğitim için dinamik maskeli dil modellemesi (MLM) kullanır ve böylece NLP problemlerinde daha başarılı sonuçlar elde eder. Metin sınıflandırma, duygu analizi, adlandırılmış varlık tanıma (NER), özetleme gibi alanlarda sıklıkla kullanılır. Şekil 2.5'te RoBERTa mimarisinin çalışma mantığı verilmiştir.



Şekil 2.5 RoBERTa mimarisinin çalışma mantığı (Mitri vd., 2024)

2.1.3.3 DistilBERT (Distilled BERT)

DistilBERT modeli BERT mimarisinin küçültülmüş ve daha hızlandırılmış bir versiyonudur. Daha büyük bir modelin davranışını yeniden üretmek üzere daha küçük bir model eğitilerek bilgi damıtması yoluyla oluşturulmuştur (Perikos ve Diamantopoulos, 2024). Model boyutu küçültülüp model hızı artırılarak BERT modeline göre daha avantajlı sonuçlar elde edilmesi amaçlanmıştır. BERT'in girdilerini

sıkıştırıp önemli olanları kullanarak eğitim yapar ve sonuca gider. Bunu yaparken daha az bellek tüketimi ile de avantaj sağlar. Duygu analizi, metin sınıflandırma, soru-cevap ve doğal dil anlama alanlarında başarılı sonuçlar verir. Özellikle kaynak kullanımının sınırlı ve zamanın görece az olduğu çalışmalarda eğitim için hesaplama gücü adına diğer modellerden daha hızlı çalışabilmesi gibi özellikleri sayesinde tercih edilmekte ve bu çalışmalarda başarılı performanslar sergilemektedir (Perikos ve Diamantopoulos, 2024).

2.1.3.4 ALBERT (A Lite BERT)

ALBERT modeli Google tarafından geliştirilmiş ve BERT modeline göre hem hız açısından hem verimlilik açısından hem de boyut bağlamında daha avantajlı bir Transformer modelidir. Kullanılan parametre sayısını azaltarak bellek kullanımını azaltmış ve daha iyi sonuçlar elde etmeyi amaçlamıştır.

Uzun metinlere sahip veri setlerinde daha iyi anlam ilişkileri kurmak ve cümle sırası tahmini yapabilmek gibi avantajları dolayısıyla ALBERT modeli başarı oranlarını artırmıştır. Büyük veri setlerinde iyi sonuçlar vermesi bu tür çalışmalarda kullanılmasına yol açmıştır. Hukuk, finans ve tıp alanlarında metin analizi, genel çalışmalarda duygu analizi, metin anlamlandırma gibi birçok işi başarıyla yapabilmektedir. ALBERT, büyük ölçekli yapılandırmalara olanak tanıyan, önceki bellek sınırlamalarını aşan ve model bozulması ile ilgili olarak iyi performans sergileyen parametre azaltma teknikleri kullanmaktadır (Perikos ve Diamantopoulos, 2024).

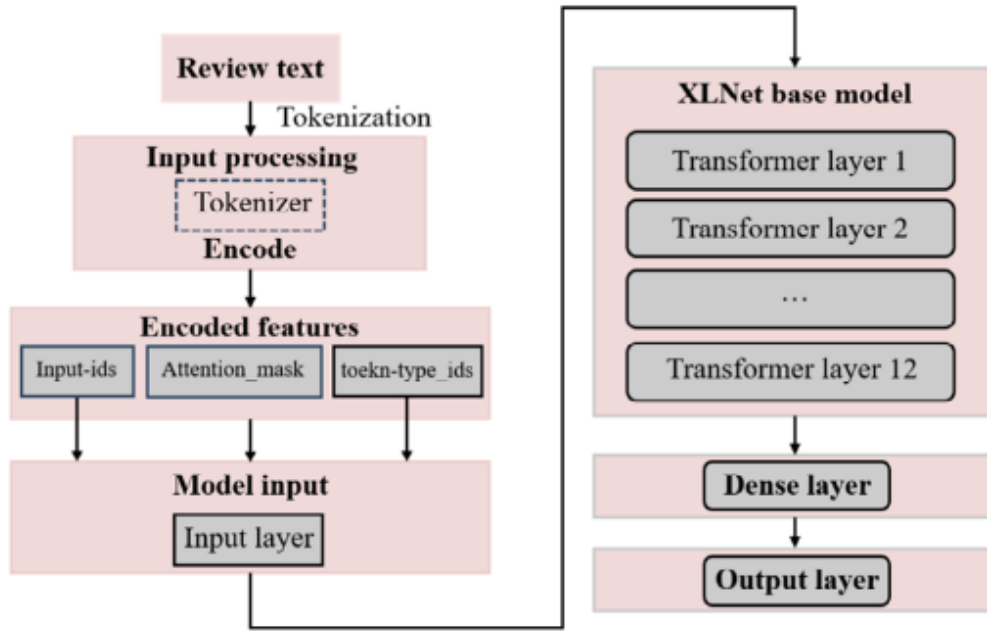
2.1.3.5 DeBERTa (Decoding-enhanced BERT with Disentangled Attention)

Microsoft tarafından NLP görevlerinde üstün başarı eldesi için BERT ve RoBERTa modelleri ilerletilerek geliştirilmiştir. Giriş katmanındaki her kelimenin, kelime (içerik) gömme ve konum gömmesinin toplamı olan bir vektör kullanılarak temsil edildiği BERT'den farklı olarak, DeBERTa'daki her kelime, sırasıyla içeriğini ve konumunu kodlayan iki vektör kullanılarak temsil edilir ve kelimeler arasındaki dikkat ağırlıkları içeriklerine ve göreceli konumlarına göre çözülmüş matrisler kullanılarak

hesaplanır (He vd., 2020). Bu yüzden model daha anlamlı bağlam ilişkileri kurar ve uzun cümlelerde sağladığı bu performans, modeli daha başarılı bir hale getirir. DeBERTa duygu analizi, metin sınıflandırma, adlandırılmış varlık tanıma gibi NLP görevlerinde başarılı sonuçlar verir.

2.1.3.6 XLNet

Google ve Carnegie Mellon University tarafından 2019 yılında geliştirilen XLNet modeli duygu analizi, metin sınıflandırma, dil modelleme gibi NLP görevlerinde oldukça başarılı sonuçlar vermeye başlamıştır. BERT modeli gibi çift sıralı bağlamı kullanmasına rağmen maskeleye yapmayıp kelimeleri permütasyon mantığında sınıflandırıp farklı sıralarda tahmin etme yöntemi kullandığı için daha başarılı bir performansa kavuşmuştur. Şekil 2.6'da XLNet mimarisinin çalışma mantığı sunulmuştur.



Şekil 2.6 XLNet mimarisinin çalışma mantığı (Shetty vd., 2025)

XLNet gizli kelimeler arasındaki bağlantıları yeterince kontrol etmek için BERT eğitim yaklaşımının temel sınırlamalarını ele almaktadır (Shetty vd., 2025). Bu haliyle

BERT mimarisinin bazı sınırlamalarını geliştirmeye odaklanmıştır. XLNet otomatik regresif modellemeyi dahil ederek bu dezavantajın üstesinden gelir (Shetty vd., 2025). Dizi uzunluğu sınırının olmaması, çift yönlü bağlamı permütasyonla çözmesi, otoregresif özelliğiyle de BERT ve türetilmiş versiyonları geride bırakabilen bir model olma yolunda ilerlemektedir.

2.1.3.7 GPT (Generative Pre-trained Transformer)

GPT modeli OpenAI tarafından geliştirilmiş, Transformer mimarisine dayalı bir dil modelidir. NLP görevlerini başarılı bir performansla yerine getiren GPT modelleri otomatik regresif modelleme yöntemini kullanırlar. Böylece kelimeleri sırası ile takip eder ve bağlamı ona göre kurarlar. MLM kullanmayan bu modeller ince ayar yapılarak birçok değişik görevde kullanılabilirler. Duygu analizi çalışmalarında da başarılı sonuçlar verirler.

GPT, verilen bir girdi içindeki bağlamı anlayarak insan benzeri metinler işlemek ve üretmek için tasarlanmış, dönüştürücü mimari üzerine inşa edilmiş büyük bir dil modelidir (Perikos ve Diamantopoulos, 2024). GPT mimarisi BERT gibi mimarilerden farklı olarak cümleleri soldan sağa doğru işler ve önceki bağlamı kullanır.

2.1.3.8 ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)

Google Research ve Stanford Üniversitesi tarafından 2020 yılında geliştirilen bu model nispeten daha az kaynak tüketimi ile eğitim yapan ve daha başarılı ve verimli sonuçlar elde etmeyi amaçlayan bir Transformer modelidir. Generatör (Üreteç) kullanarak MLM ile az sayıda kelimeyi maskeler ve daha sonra discriminator (ayrıt edici) ile de bu kelimelerin sahte ya da doğru olduğunu saptamaya çalışır. Bu modelin amacı, giriş belirteçlerini küçük bir transformatör ağı tarafından üretilen yüksek kaliteli negatif örneklerden ayırt etmek için bir metin kodlayıcı eğitmektir (Güven, 2022). BERT modeline göre ön eğitimde daha az hesaplama gücü kullanır. NLP işlemlerinin hemen hepsinde başarılı bir şekilde kullanılır.

2.1.3.9 T5 (Text-to-Text Transfer Transformer)

2019 yılında Google Research tarafından üretilen T5 mimarisi text-to-text yani metinden metne dönüşüm yapan ve böylece bütün NLP görevlerinde başarılı sonuçlar üreten bir Transformer mimarisidir. Denetimli ve denetimsiz veri kümelerinin bir karışımı üzerinde eğitilen T5, diğer Transformatör tabanlı modellere benzer şekilde kodlayıcı-kod çözücü Transformatör mimarisi aracılığıyla çalışır (Shaik Vadla vd., 2024). Farklı boyutlarda modeller sunabilme imkanına sahip bu Transformer mimarisi T5-Small, T5-Base, T5-Large, T5-3B (3 milyar parametrelili) ve T5-11B (11 milyar parametrelili) boyutlardaki modelleriyle küçük büyük bütün veri setlerinde başarılı bir çalışma imkanına sahiptir. Bu geniş ön eğitimi sayesinde bütün NLP problemleri için uyarlanabilip dönüştürülebilme yeteneğine sahiptir ve ince ayar yapılarak da birçok alanda kullanılabilmektedir.

T5'in tıpkı bir madeni para gibi iki yüzü vardır; kodlayıcı metninizi analiz eder, anlamını ve kelimeler arasındaki ilişkileri anlar (Shaik Vadla vd., 2024). Daha sonra kod çözücü tarafından oluşturulan yanıtlar ile modelin kullanılmak istendiği alanda kelimeler bir araya getirilir.

2.1.3.10 BART (Bidirectional and Auto-Regressive Transformer)

Meta AI tarafından 2019'da geliştirilen BART modeli BERT ve GPT modellerinin özelliklerinin harmanlanarak oluşturulduğu bir Transformer modelidir. BART, BERT modelinin kodlayıcı (encoder) özelliği ile GPT modelinin çözücü (decoder) özelliğini birleştirerek hem bağlamsal anlama problemlerinde hem de metin vb. üretimi konularında başarı sağlamıştır.

BART'ın diğer önceden eğitilmiş modellere göre avantajı, otoregresif doğası nedeniyle BERT gibi davranan 6 katmanlı bir kodlayıcı ve GPT gibi davranan 6 katmanlı bir kod çözücünün yalnızca önceden eğitilmiş 12 katmanlı kodlayıcı modeli BERT'in yerini alabilmesidir (Suresh ve Banu, 2024). Böylece hem metinleri çift yönlü işleyebilme hem de soldan sağa doğru metin üretme işlevlerini bir arada yapabilmektedir. BART ayrıca bozulmuş metinleri düzeltme ve gürültü giderme özelliklerine de sahiptir.

2.2 Accuracy (Doğruluk)

Doğruluk modelin doğru tahmin etme oranını ölçen bir metriktir. Doğru bir şekilde tahmin edilen örneklerin toplam örnek sayısına oranıdır. Aşağıdaki formülle ifade edilir.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1)$$

TP (True Positive): Gerçekte pozitif olup modelin pozitif olarak tahmin ettiği değerler.

TN (True Negative): Gerçekte negatif olup modelin negatif olarak tahmin ettiği değerler.

FP (False Positive): Gerçekte negatif olup modelin pozitif olarak tahmin ettiği değerler.

FN (False Negative): Gerçekte pozitif olup modelin negatif olarak tahmin ettiği değerler.

2.3 Classification Report

Sınıflandırma problemlerinde performans değerlendirmesi için kullanılan bir rapordur. Precision, Recall, F1-score, Support gibi metrikleri barındırır. Scikit-learn kütüphanesinden model içerisine aktarılıp kullanılır. Çizelge 2.1’de bir örneği gösterilmiştir.

Çizelge 2.1 Classification Report örneği

Class	Precision	Recall	F1-Score	Support
Negative	0.86	0.91	0.89	4427
Neutral	0.90	0.83	0.86	3678
Positive	0.87	0.88	0.88	4120
Accuracy			0.88	12225
Macro Avg	0.88	0.87	0.88	12225
Weighted Avg	0.88	0.88	0.88	12225

2.4 Precision (Kesinlik)

Modelin tahmin ettiği doğru pozitif örneklerin bütün pozitif örneklere oranıdır. Bir modelin tahmin ettiği pozitif sonuçların gerçekte ne kadarının pozitif olduğudur. Aşağıdaki formülle ifade edilir.

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

2.5 Recall (Duyarlılık)

Modelin gerçek pozitif örnekleri hangi oranda yakaladığını gösteren bir değerlendirme metriğidir. Modelin yakalayamadığı pozitif örnekleri yanlış negatif örnekleri göz önünde bulundurarak ölçer.

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

Recall değerinin yüksek olması modelin pozitif değerleri iyi yakalayabildiğini fakat yanlış pozitif örneklerin de fazla olabileceğini gösterir.

2.6 F1-Score

Sınıflandırma problemlerinde model performansını ölçen bir metriktir. F1-score precision ve recall metriklerinin harmonik ortalamasıdır. Dengesiz veri setlerinde precision ve recall metrikleri arasındaki dengeyi sağlar ve accuracy metriği yerine daha anlamlı olarak kullanılabilir.

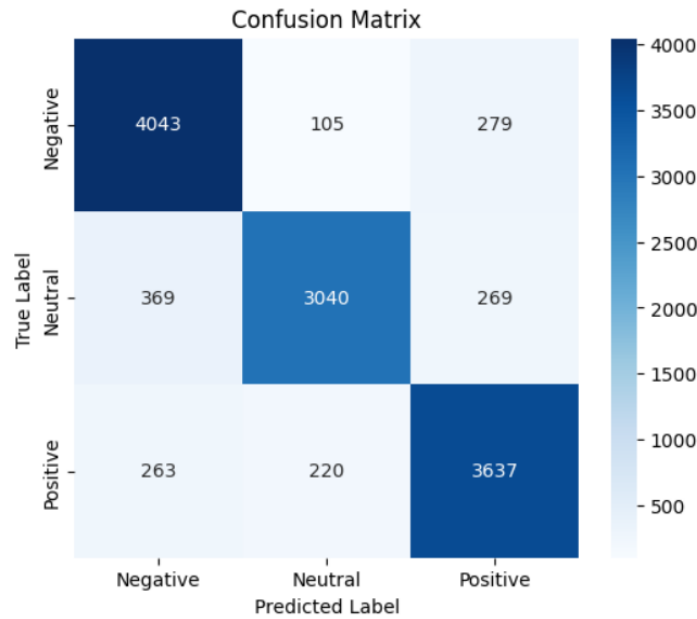
$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

2.7 Support

Bir sınıfa ait modelde kullanılan gerçek örneklem sayısını verir.

2.8 Confusion Matrix (Karmaşıklık Matrisi)

Karmaşıklık matrisi modelin tahmin ettiği örneklerle gerçek örnekleri karşılaştıran bir değerlendirme tablosudur. Hem gerçek doğru ve yanlış örnekleri hem de modelin tahmin yaptığı doğru ve yanlış örnekleri bir arada gösterir (Şekil 2.7).



Şekil 2.7 Confusion Matrix örneği



3. MATERYAL VE YÖNTEM

Bu bölümde tez çalışmasında kullanılan materyal ve uygulanan yöntemler açıklanmıştır.

3.1 Veri Seti

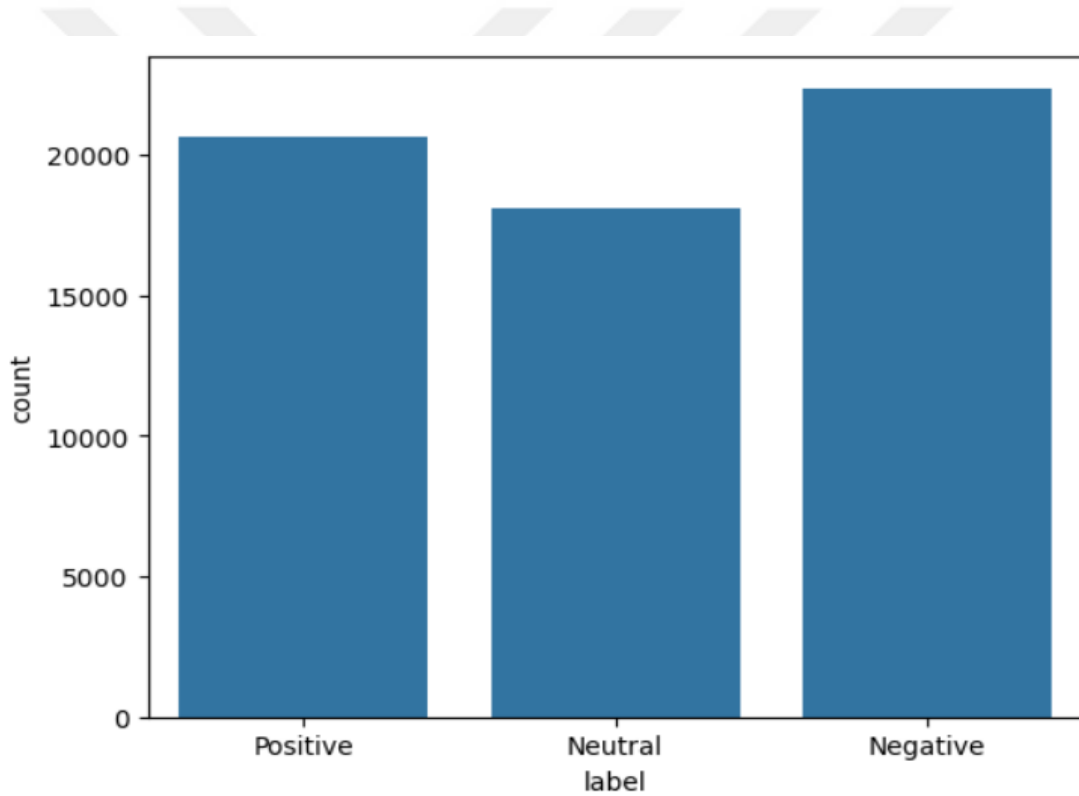
Çalışmada veri seti olarak Kaggle platformunda açık kaynak olarak paylaşılan Twitter (X) yorumlarından oluşan bir veri seti ve yine Kaggle platformundan alınan Emotion adlı üç ayrı duygu sınıfından oluşan bir veri seti kullanılmıştır.

3.1.1 Twitter (X) Veri Seti

Twitter (X) veri setinde benzersiz bir kimlik tanımlaması olan id numaraları sütunu, yorumların nereye yapıldığıyla alakalı bir başlık(topic) sütunu, yorum duygu sınıflandırması için pozitif, negatif, nötr ve alakasız etiketlerinden oluşan bir label sütunu ve yorum metinlerinden oluşan text sütunu bulunmaktadır. Çalışmada label ve text sütunları kullanılmıştır. Ayrıca alakasız etiketli satırlar çıkarılmış olup çalışmanın daha iyi sonuç elde etmesi amaçlanmıştır. Veri seti için gerekli ön hazırlıklar yapılmış olup veri makine dili için en uygun hale getirilmiştir. Bunun için eksik veri temizlemesi, boş satırların kaldırılması, verilerin sayısallaştırılması gibi temel NLP adımları uygulanmıştır. Ham halde 74682 satırdan oluşan veri seti temizlik ve ön hazırlık çalışmalarından sonra 61121 satır olarak işlenmiştir. Çizelge 3.1'de Twitter (X) veri setindeki ilk beş satır gösterimi ve Şekil 3.1'de ise Twitter (X) veri setindeki label sütununun örnek sayıları sunulmuştur.

Çizelge 3.1 Twitter (x) veri setindeki ilk beş satır gösterimi

	label	text
0	Positive	im getting on borderlands and i will murder you all ,
1	Positive	I am coming to the borders and I will kill you all,
2	Positive	im getting on borderlands and i will kill you all,
3	Positive	im coming on borderlands and i will murder you all,
4	Positive	im getting on borderlands 2 and i will murder you me all,



Şekil 3.1 Twitter (x) veri setindeki label sütununun örnek sayıları

Kelime bulutu, bir metin içerisinde en sık kullanılan kelimelerin görsel olarak öne çıkarıldığı bir analiz ve sunum aracıdır. Bu yöntemde kelimeler metinde geçme sıklıklarına göre farklı büyüklüklerde gösterilir; en sık geçen kelimeler daha büyük, daha az geçen kelimeler ise daha küçük puntolarla yer alır. Kelime bulutları metindeki temel kavramları hızlı bir şekilde görselleştirmek amacıyla sıklıkla kullanılır. Sosyal



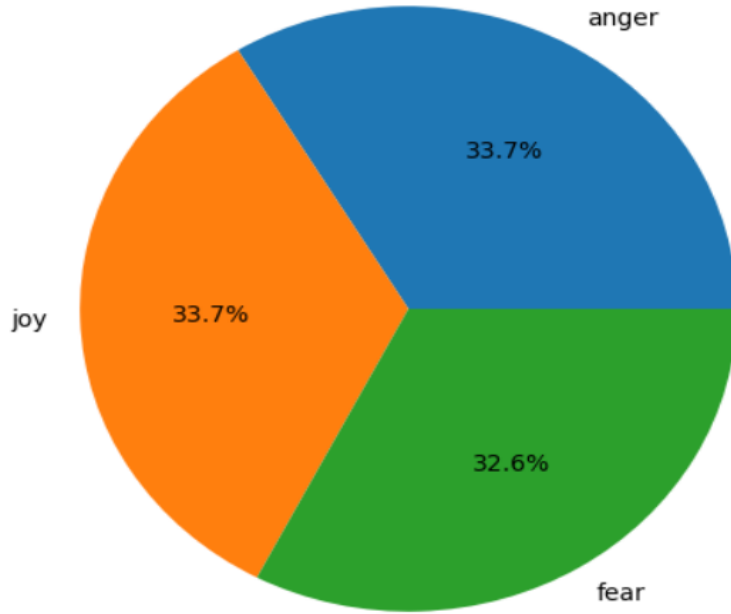
3.1.2 Emotion Veri Seti

Kaggle platformundan alınan Emotion veri seti Doğal Dil İşleme ve duygu analizi konularında çalışmaları kolaylaştırmak için oluşturulmuştur. Anger (kızgın), joy (neşeli) ve fear (korkmuş) duygu sınıflarından oluşan bu veri seti 5937 satırdan oluşmaktadır. Bu veri seti duygu sınıflandırması için hazırlanmış ve gerekli ön adımlar uygulanmıştır. Çizelge 3.2 Emotion veri setindeki ilk beş satır gösterimini içermektedir.

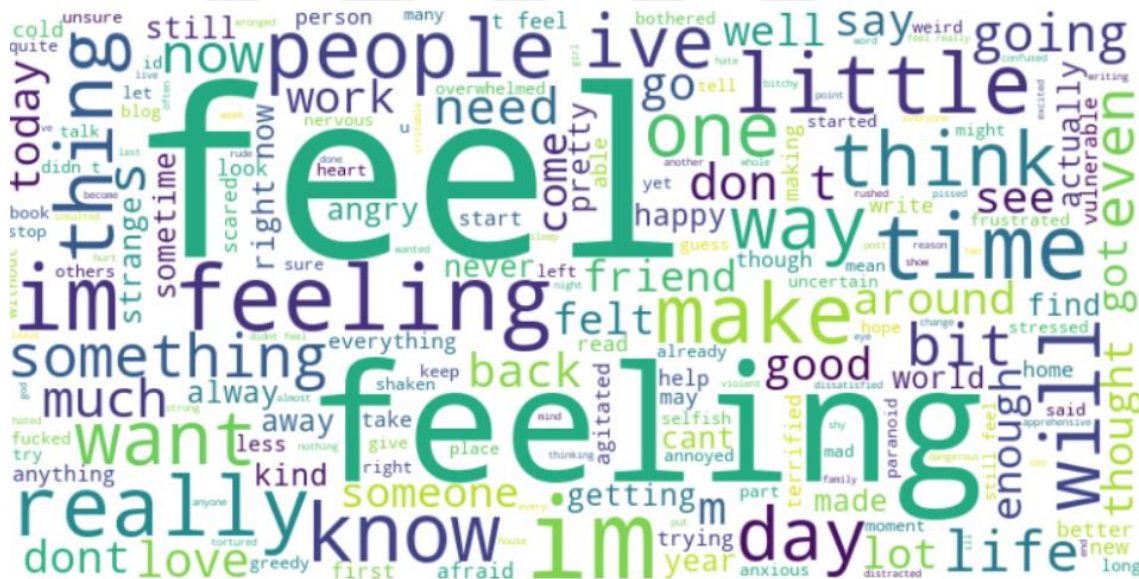
Çizelge 3.2 Emotion veri setindeki ilk beş satır gösterimi

	Comment	Emotion
0	i seriously hate one subject to death but now ...	fear
1	im so full of life i feel appalled	anger
2	i sit here to write i start to dig out my feel...	fear
3	ive been really angry with r and i feel like a...	joy
4	i feel suspicious if there is no one outside l...	fear

Emotion veri setinde boş satır bulunmamakta olup veri etiket sayıları birbirine çok yakındır. Etiket dağılımı Şekil 3.4'te ve temizlik ve ön işleme adımlarından önceki ve sonraki kelime bulutları da Şekil 3.5 ve 3.6'da gösterilmiştir.



Şekil 3.4 Emotion veri setindeki Emotion sütununun etiket dağılımı



3.2 Python

Python programlama dili 1991 yılında Guido van Rossum tarafından geliştirilen hızlı ve dinamik bir dildir. Basit ve okunaklı söz dizimine sahip olması, yorumlanabilir olması, geniş bir kütüphane yelpazesi sunması gibi avantajlarıyla öne çıkmıştır.

Python veri bilimi, makine öğrenmesi, yapay zeka, çeşitli alanlarda oyun programlama ve web programlama gibi birçok çeşitli alanda yaygın olarak kullanılır. Barındırdığı çok sayıda kütüphane ile bu alanlarda tercih edilen dil özelliğine kavuşmuştur.

Bu tez çalışmasında yazılan kodlar Python programlama dili ile yazılmıştır. Veri temizleme olarak bilinen yanlış verilerin kolayca kaldırılıp düzeltilebilmesi, verilerin ayıklanıp çeşitli özelliklerinin seçilebilmesi, veri etiketlemenin rahatça yapılabilmesi, verilerden farklı öngörüler çıkarılabilmesi ve verinin anlaşılmasında etkisi yüksek olan grafiklerin kullanılmasında birçok avantaj sağlaması gibi özellikleri Python dilinin veri bilimi ve yapay zeka çalışmalarında öne çıkmasını sağlamıştır (Anonim, 2022). Bu ve benzeri özellikleri dolayısıyla Python programlama dili tercih edilmiştir.

3.3 Google Colaboratory (Colab)

Google Colaboratory Google platformunun ücretsiz bir şekilde kullanıma sunduğu bir notebook hizmetidir. Özellikle yapay zeka ve veri bilimi çalışmalarında oldukça fazla tercih edilen bir platformdur. Harhangi bir yerden kod yazma özelliğine sahip bu platform ücretsiz GPU/TPU desteği, drive ile entegre edilebilir olması, Python destekli bir platform olması yönünden de çok avantajlı bir konumdadır. Bu özellikleri sebebiyle bu tez çalışmasındaki notebooklar colab platformunda çalıştırılmıştır.

3.4 Tokenizasyon

Tokenizasyon işlemi metinleri, cümleleri daha küçük parçalara bölme işlemidir. Böylece metinleri makinelerin anlayabileceği sayısal dile çevirmiş olurlar. Bu işlem sonucunda token adı verilen parçalar sayesinde NLP işlemleri çok daha rahat uygulanabilir hale gelir. Tokenizer adı verilen araçlarla bu işlemler yapılır. Tokenizerler metin verilerini sayısal verilere dönüştürerek makine diline çeviren araçlardır.

Transformer tabanlı modeller metinleri doğrudan işleyemediklerinden tokenizerlere ihtiyaç duyarlar. Tokenizerler metinleri parçalara bölüp sayı dizilerine çevirirler. Gerekğinde maskeleme yapıp metinleri makine için daha anlaşılır hale getirirler. Transformer tabanlı modellerde tokenizerler metinleri kelimelere ayırmakla

kalmaz gerektiğinde kelime köklerine kadar inip parçalama yaparlar. Subword Tokenizasyon adı verilen bu uygulama sayesinde metinler daha iyi ve verimli bir şekilde anlaşılmış olur. Bu modellerde lemmatizasyon görevini model kendisi yaptığı için lemmatizasyon gereksizdir hatta modelin eğitimine zarar bile verebilmektedir. Word2Vec ve GloVe gibi embedding yöntemlerini ise model zaten kendisi işlediği için klasik makine öğrenmesi yöntemlerinden farklı olarak bu modellerde kullanılmaz.

Tokenizasyon işlemi metni alt parçalara bölme, noktalama işaretlerini makine diline uygun bir şekilde işleme, metni sayılara yani vektörlere çevirme, padding ve truncation işlemleri yaparak sabit uzunlukta giriş sağlama, masklar oluşturma ve special token ekleme gibi işlemleri bir arada yaparak Transformer modeller için birçok avantaj sağlamaktadır. Bundan dolayı bu tez çalışmasında her bir Transformer modeli için kendi tokenizerleri kullanılmıştır.

3.5 Transformer Tabanlı Modellerin Eğitimi ve Çalıştırılma Yöntemi

Bu tez kapsamında kullanılan Transformer tabanlı modeller, literatürde yaygın şekilde kullanılan ve birbirinden farklı mimari yaklaşımları temsil eden on modelden oluşmaktadır. Model seçiminde, yalnızca yüksek başarı oranlarına sahip olan yaklaşımlar değil aynı zamanda farklı yapısal özellikler barındıran ve çeşitli veri boyutlarında ve görev türlerinde performans göstermesiyle öne çıkan modeller tercih edilmiştir.

Bu kapsamda encoder tabanlı mimariler (BERT, RoBERTa, ALBERT, DistilBERT, DeBERTa), decoder veya encoder-decoder temelli mimariler (GPT, T5, BART) ile discriminative-generative hibrit yapılar (ELECTRA, XLNet) dâhil edilmiştir. Böylece farklı mimari tasarımların duygu analizi görevlerindeki performansları nesnel bir biçimde karşılaştırılmıştır. Ayrıca bu modellerin Hugging Face Transformers kütüphanesi gibi açık kaynak araçlarla kolay erişilebilir olması, önceden eğitilmiş ağırlıklara sahip olmaları ve hem büyük hem de küçük veri kümeleri üzerinde uygulanabilir olmaları da seçim kriterleri arasında yer almıştır. Model çeşitliliği sayesinde gerçek dünya uygulamalarına yönelik farklı senaryoların değerlendirilmesine olanak tanınmıştır.

ALBERT, BART, BERT, DeBERTa, DistilBERT, ELECTRA, GPT, RoBERTa, T5 ve XLNet Transformer modelleri kullanılarak gerçekleştirilen bu çalışmada bütün notebooklar için veri temizleme ve ön işleme adımları uygulanmıştır. Transformer tabanlı modellerde tokenizerler lemmatizasyon ve kelime gömme gibi adımları kendileri yaptığından bu işlemler yapılmamıştır. Bu adımlar Transformer tabanlı modellerde önceden eğitilmiş token dizilerini bozmak ve verinin gerçek anlamını azaltmak gibi sonuçlara yol açabilmektedirler. Bu da Transformer modelleri için başarı oranının azalmasına yol açabilmektedir. Aynı şekilde Transformer tabanlı modeller kelimeleri subwordlerine yani bir alt kelimelerine ayırdığı için tekrar embedding (kelime gömme) işlemi uygulamak mantıksızdır. Bu modeller için ön adımların sadeleştirilmesi modellerin başarı oranını artırmaktadır.

Temizlik ve sadeleştirme gibi ön işleme adımlarından sonra her bir model için ilgili tokenizeri çağırılmış ve kullanılmıştır. Modellerin her biri 5 epoch boyunca çalıştırılmış ve accuracy, precision, recall, F1-score gibi değerlendirme metrikleriyle değerlendirilmiştir. Ayrıca karmaşıklık matrisi ve kayıp grafikleri ile de sonuçlar analiz edilmiştir.



4. BULGULAR

Bu bölümde Transformer tabanlı modellerin eğitimi için çalıştırılan notebookların sonuçları ve çeşitli metriklerle değerlendirilmeleri yapılmıştır. On ayrı Transformer modeli için colab ortamında iki ayrı veri seti ile birer tane notebook çalıştırılmıştır. Her bir model için beş epoch ile eğitim yapılmıştır. Eğitimlerin sonuçları tablolar ve grafikler halinde verilmiştir.

Bütün modeller kendi özel tokenizerleriyle çalıştırılmıştır. T5 modeli 8 batch_size ve diğer modeller 16 batch_size değerleriyle eğitilmiştir. Öğrenme oranları içinse BART, BERT, DistilBERT, GPT, T5 ve XLNet modelleri 5×10^{-5} learning_rate değeri ile ve ALBERT, DeBERTa, ELECTRA ve RoBERTa modelleri 2×10^{-5} learning_rate değeri ile çalıştırılmıştır.

4.1 Kullanılan Modeller

Bu çalışma kapsamında ele alınan modeller ve alınan sonuçlar aşağıda açıklanmıştır.

4.1.1 BERT

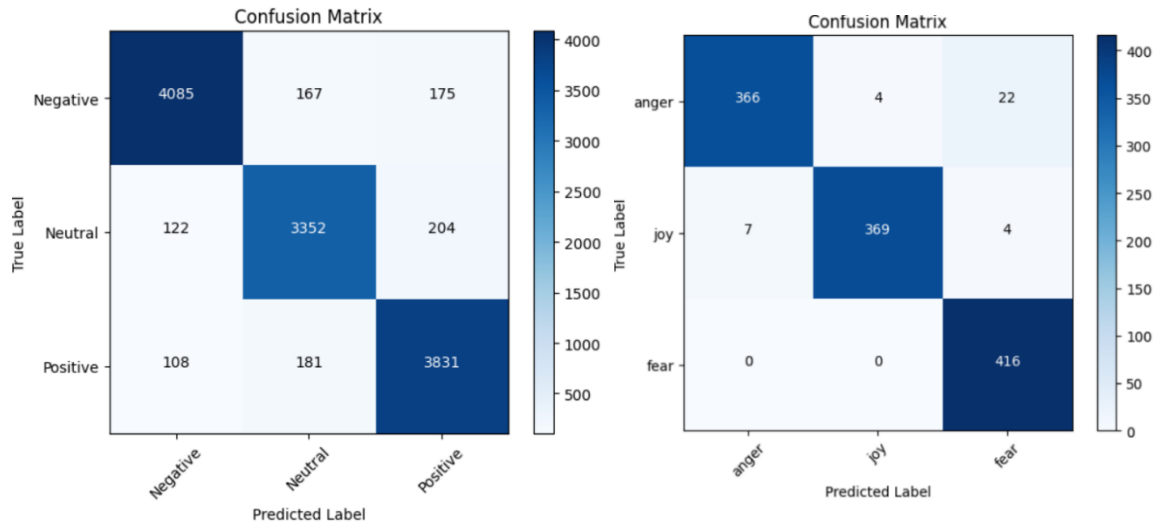
BERT modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile % 92.17 ve Emotion veri seti ile % 96.88 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.1, Çizelge 4.2 ve Şekil 4.1’de BERT modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.1 BERT modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.95	0.92	0.93	4427
Neutral	0.91	0.91	0.91	3678
Positive	0.91	0.93	0.92	4120
Accuracy			0.92	12225
Macro Avg	0.92	0.92	0.92	12225
Weighted Avg	0.92	0.92	0.92	12225

Çizelge 4.2 BERT modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.98	0.93	0.96	392
joy	0.99	0.97	0.98	380
fear	0.94	1.00	0.97	416
accuracy			0.97	1188
macro avg	0.97	0.97	0.97	1188
weighted avg	0.97	0.97	0.97	1188



Şekil 4.1 BERT modeli için karmaşıklık matrisleri

4.1.2 DistilBERT

DistilBERT modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile % 91.78 ve Emotion veri seti ile % 97.05 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.3 ve Çizelge 4.4'te DistilBERT modeli için alınan sonuçlar tablolar halinde sunulmuştur.

Çizelge 4.3 DistilBERT modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.89	0.96	0.92	4427
Neutral	0.93	0.90	0.91	3678
Positive	0.95	0.89	0.92	4120
accuracy			0.92	12225
macro avg	0.92	0.92	0.92	12225
weighted avg	0.92	0.92	0.92	12225

Çizelge 4.4 DistilBERT modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.96	0.97	0.96	392
joy	0.99	0.97	0.98	380
fear	0.97	0.98	0.97	416
accuracy			0.97	1188
macro avg	0.97	0.97	0.97	1188
weighted avg	0.97	0.97	0.97	1188

4.1.3 DeBERTa

DeBERTa modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile % 90.96 ve Emotion veri seti ile % 98.06 oranında accuracy (doğruluk)

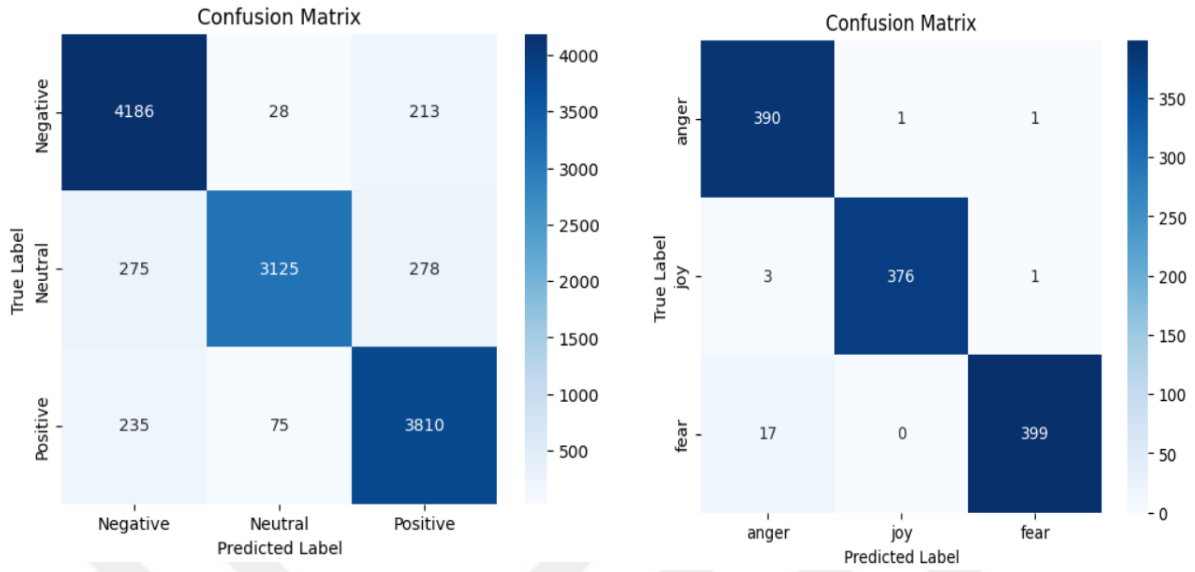
değerleri elde edilmiştir. Çizelge 4.5, Çizelge 4.6 ve Şekil 4.2’de DeBERTa modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.5 DeBERTa modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.89	0.95	0.92	4427
Neutral	0.97	0.85	0.91	3678
Positive	0.89	0.92	0.90	4120
accuracy			0.91	12225
macro avg	0.92	0.91	0.91	12225
weighted avg	0.91	0.91	0.91	12225

Çizelge 4.6 DeBERTa modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.95	0.99	0.97	392
joy	1.00	0.99	0.99	380
fear	1.00	0.96	0.98	416
accuracy			0.98	1188
macro avg	0.98	0.98	0.98	1188
weighted avg	0.98	0.98	0.98	1188



Şekil 4.2 DeBERTa modeli için karmaşıklık matrisleri

4.1.4 RoBERTa

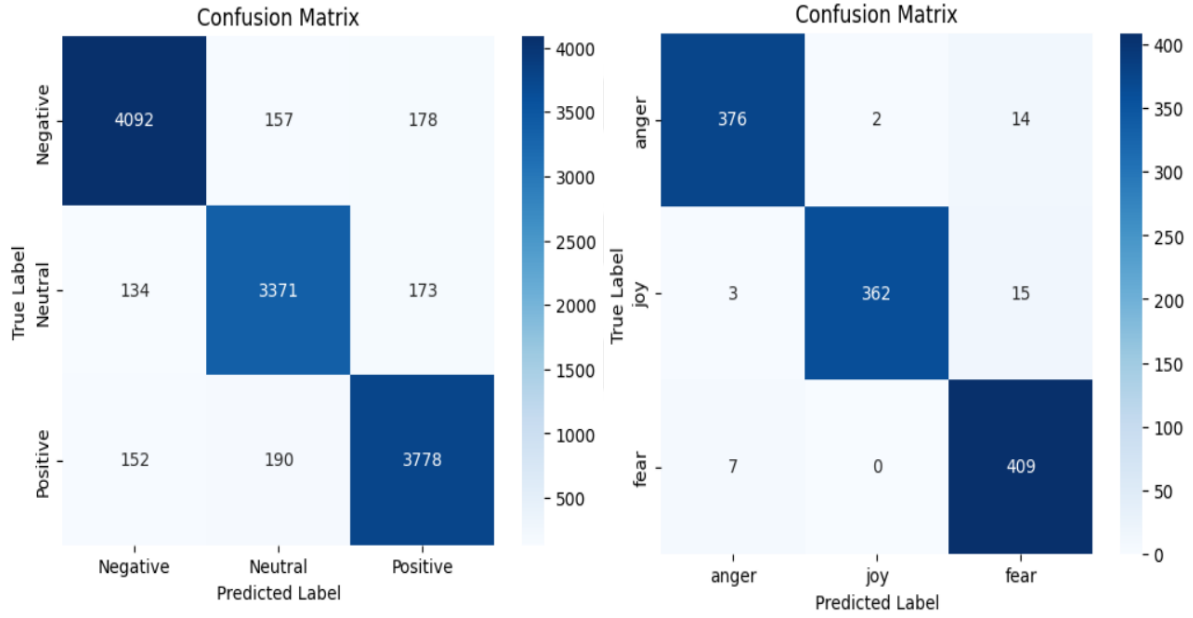
RoBERTa modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %91.56 ve Emotion veri seti ile % 96.54 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.7, Çizelge 4.8 ve Şekil 4.3'te RoBERTa modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.7 RoBERTa modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.93	0.92	0.93	4427
Neutral	0.91	0.92	0.91	3678
Positive	0.91	0.92	0.92	4120
accuracy			0.92	12225
macro avg	0.92	0.92	0.92	12225
weighted avg	0.92	0.92	0.92	12225

Çizelge 4.8 RoBERTa modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.97	0.96	0.97	392
joy	0.99	0.95	0.97	380
fear	0.93	0.98	0.96	416
accuracy			0.97	1188
macro avg	0.97	0.96	0.97	1188
weighted avg	0.97	0.97	0.97	1188



Şekil 4.3 RoBERTa modeli için karmaşıklık matrisleri

4.1.5 ALBERT

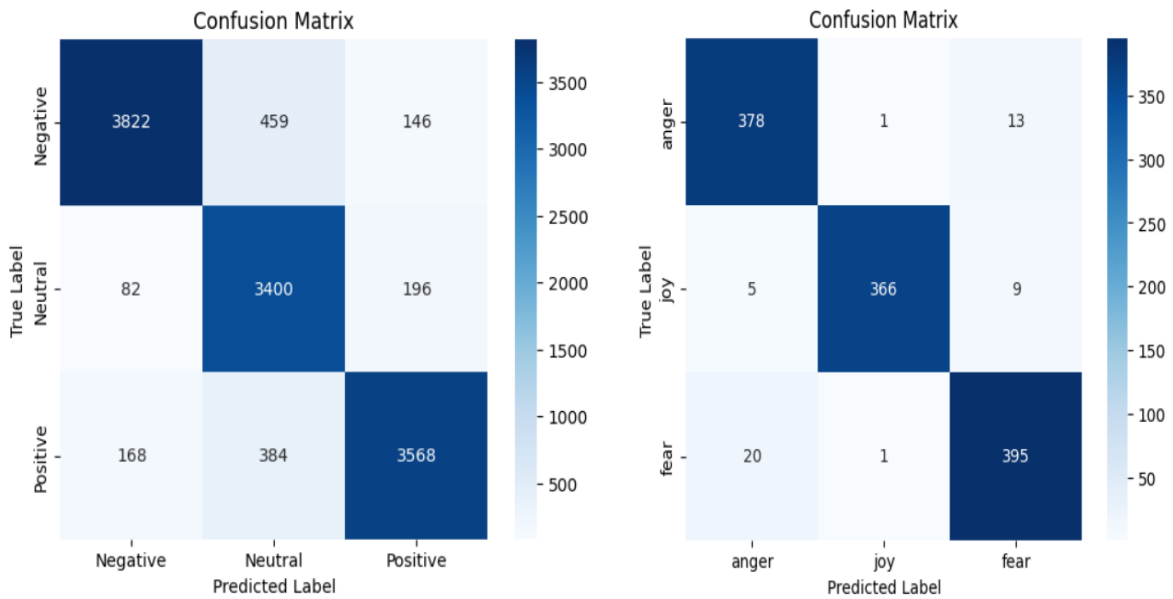
ALBERT modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %88.26 ve Emotion veri seti ile % 95.87 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.9, Çizelge 4.10 ve Şekil 4.4'te ALBERT modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.9 ALBERT modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.94	0.86	0.90	4427
Neutral	0.80	0.92	0.86	3678
Positive	0.91	0.87	0.89	4120
accuracy			0.88	12225
macro avg	0.88	0.88	0.88	12225
weighted avg	0.89	0.88	0.88	12225

Çizelge 4.10 ALBERT modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.94	0.96	0.95	392
joy	0.99	0.96	0.98	380
fear	0.95	0.96	0.95	416
accuracy			0.96	1188
macro avg	0.96	0.96	0.96	1188
weighted avg	0.96	0.96	0.96	1188



Şekil 4.4 ALBERT modeli için karmaşıklık matrisleri

4.1.6 ELECTRA

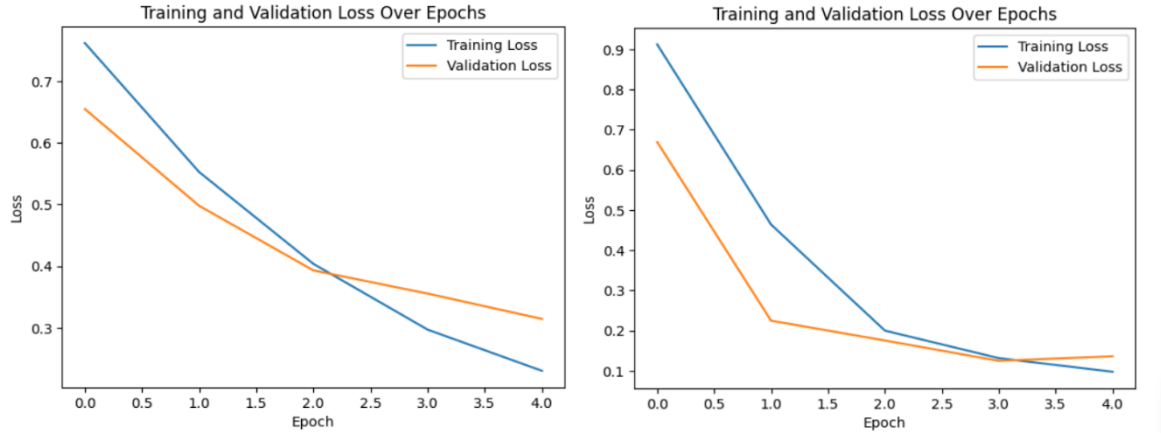
ELECTRA modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %88.78 ve Emotion veri seti ile % 95.45 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.11, Çizelge 4.12 ve Şekil 4.5'te ELECTRA modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.11 ELECTRA modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.88	0.92	0.90	4427
Neutral	0.89	0.85	0.87	3678
Positive	0.90	0.88	0.89	4120
accuracy			0.89	12225
macro avg	0.89	0.89	0.89	12225
weighted avg	0.89	0.89	0.89	12225

Çizelge 4.12 ELECTRA modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.90	0.98	0.94	392
joy	1.00	0.96	0.98	380
fear	0.97	0.93	0.95	416
accuracy			0.95	1188
macro avg	0.96	0.96	0.96	1188
weighted avg	0.96	0.95	0.96	1188



Şekil 4.5 ELECTRA modeli için eğitim ve validasyon kayıp grafikleri

4.1.7 GPT

GPT modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %91.45 ve Emotion veri seti ile % 97.05 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.13 ve Çizelge 4.14’te GPT modeli için alınan sonuçlar tablolar halinde sunulmuştur.

Çizelge 4.13 GPT modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.95	0.91	0.93	4427
Neutral	0.87	0.94	0.90	3678
Positive	0.92	0.90	0.91	4120
accuracy			0.91	12225
macro avg	0.91	0.92	0.91	12225
weighted avg	0.92	0.91	0.91	12225

Çizelge 4.14 GPT modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.93	0.98	0.96	392
joy	0.99	0.99	0.99	380
fear	0.98	0.94	0.96	416
accuracy			0.97	1188
macro avg	0.97	0.97	0.97	1188
weighted avg	0.97	0.97	0.97	1188

4.1.8 T5

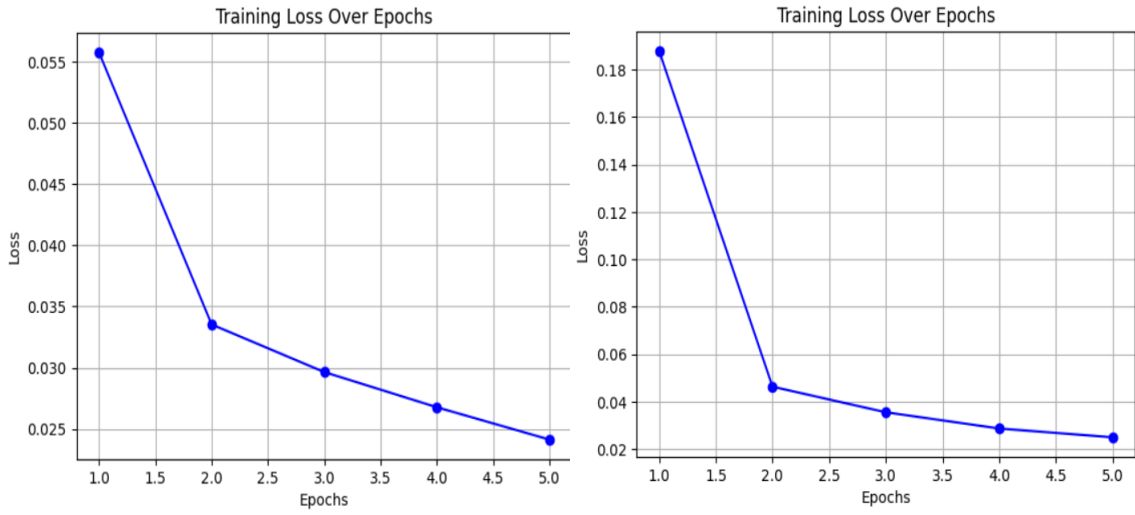
T5 modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %76.01 ve Emotion veri seti ile %89.47 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.15, Çizelge 4.16 ve Şekil 4.6’da T5 modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.15 T5 modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.78	0.83	0.80	4452
Neutral	0.76	0.62	0.69	3598
Positive	0.74	0.80	0.77	4117
accuracy			0.76	12167
macro avg	0.76	0.75	0.75	12167
weighted avg	0.76	0.76	0.76	12167

Çizelge 4.16 T5 modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.89	0.86	0.88	400
joy	0.95	0.90	0.92	400
fear	0.84	0.93	0.88	388
accuracy			0.89	1188
macro avg	0.90	0.90	0.90	1188
weighted avg	0.90	0.89	0.90	1188



Şekil 4.6 T5 modeli için epoch başına ortalama kayıp grafikleri

4.1.9 BART

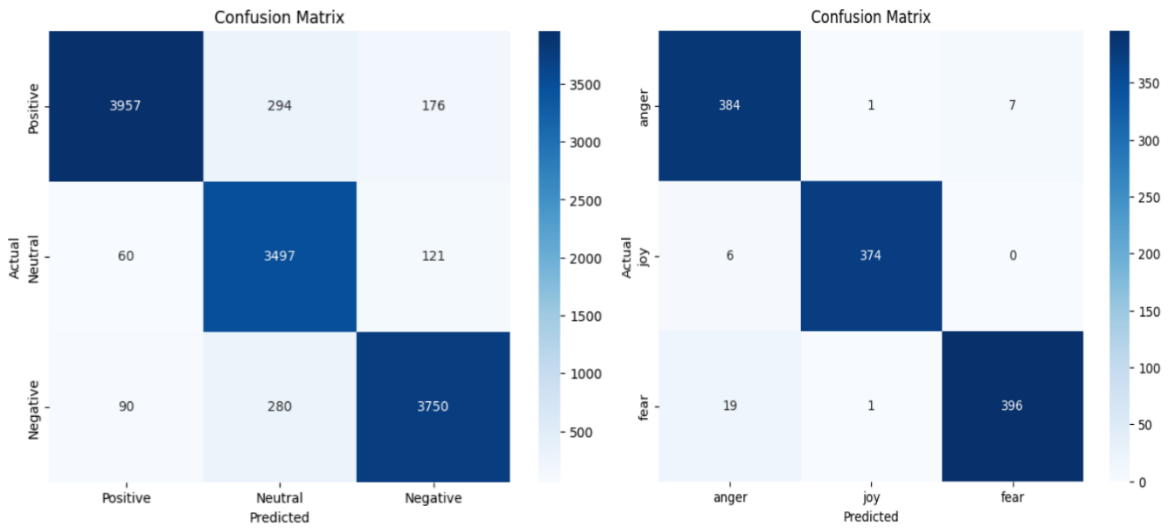
BART modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %91.65 ve Emotion veri seti ile % 97.14 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.17, Çizelge 4.18 ve Şekil 4.7’de BART modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.17 BART modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.96	0.89	0.93	4427
Neutral	0.86	0.95	0.90	3678
Positive	0.93	0.91	0.92	4120
accuracy			0.92	12225
macro avg	0.92	0.92	0.92	12225
weighted avg	0.92	0.92	0.92	12225

Çizelge 4.18 BART modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.94	0.98	0.96	392
joy	0.99	0.98	0.99	380
fear	0.98	0.95	0.97	416
accuracy			0.97	1188
macro avg	0.97	0.97	0.97	1188
weighted avg	0.97	0.97	0.97	1188



Şekil 4.7 BART modeli için karmaşıklık matrisleri

4.1.10 XLNet

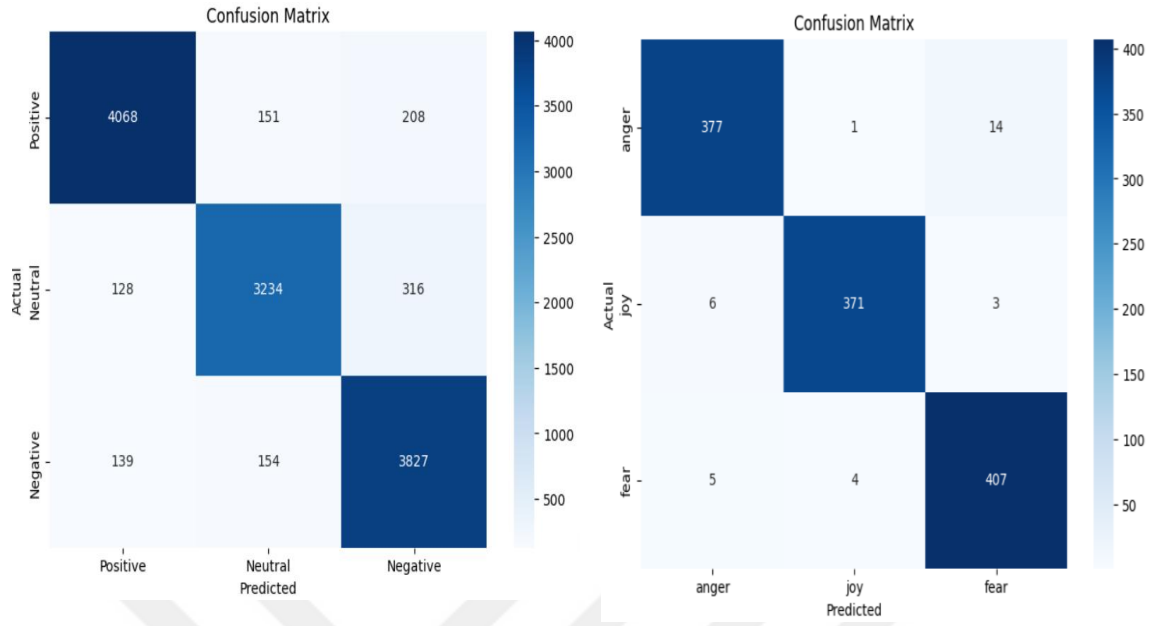
XLNet modeli için Twitter (x) ve Emotion veri setleri ile beşer epochtan oluşan birer tane eğitim döngüsü oluşturulmuştur. Bu eğitimler sonucunda Twitter (x) veri seti ile %91.03 ve Emotion veri seti ile % 97.22 oranında accuracy (doğruluk) değerleri elde edilmiştir. Çizelge 4.19, Çizelge 4.20 ve Şekil 4.8’de XLNet modeli için alınan sonuçlar tablolar ve görseller halinde sunulmuştur.

Çizelge 4.19 XLNet modeli için Twitter (x) veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
Negative	0.94	0.92	0.93	4427
Neutral	0.91	0.88	0.90	3678
Positive	0.88	0.93	0.90	4120
accuracy			0.91	12225
macro avg	0.91	0.91	0.91	12225
weighted avg	0.91	0.91	0.91	12225

Çizelge 4.20 XLNet modeli için Emotion veri seti ile alınan sonuçlar

Class	Precision	Recall	F1-Score	Support
anger	0.97	0.96	0.97	392
joy	0.99	0.98	0.98	380
fear	0.96	0.98	0.97	416
accuracy			0.97	1188
macro avg	0.97	0.97	0.97	1188
weighted avg	0.97	0.97	0.97	1188



Şekil 4.8 XLNet modeli için karmaşıklık matrisleri

4.2 Değerlendirme

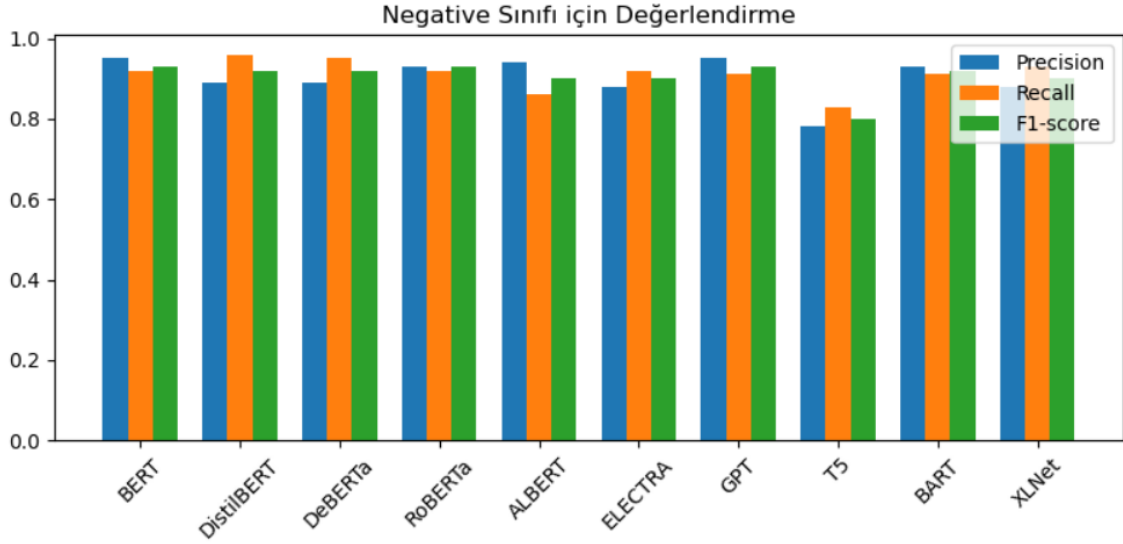
Tez çalışması kapsamında yapılan Transformer model tabanlı çalışmalarda Twitter (x) veri setinde BERT modeli %92.17 doğruluk değeri ile en yüksek doğruluk oranını yakalamış ve Emotion veri setinde %98.06 doğruluk oranı ile DeBERTa modeli en yüksek doğruluk oranına ulaşmıştır. Twitter (x) veri seti için en düşük başarı oranına %76.01 ile T5 modeli ulaşmış ve Emotion veri seti için %89.47 ile yine en düşük başarıyı T5 modeli yakalamıştır.

4.2.1 Twitter (x) Veri Seti

Twitter (X) veri seti ile yapılan çalışmalarda çeşitli değerlendirme metrikleri için şu sonuçlar alınmıştır:

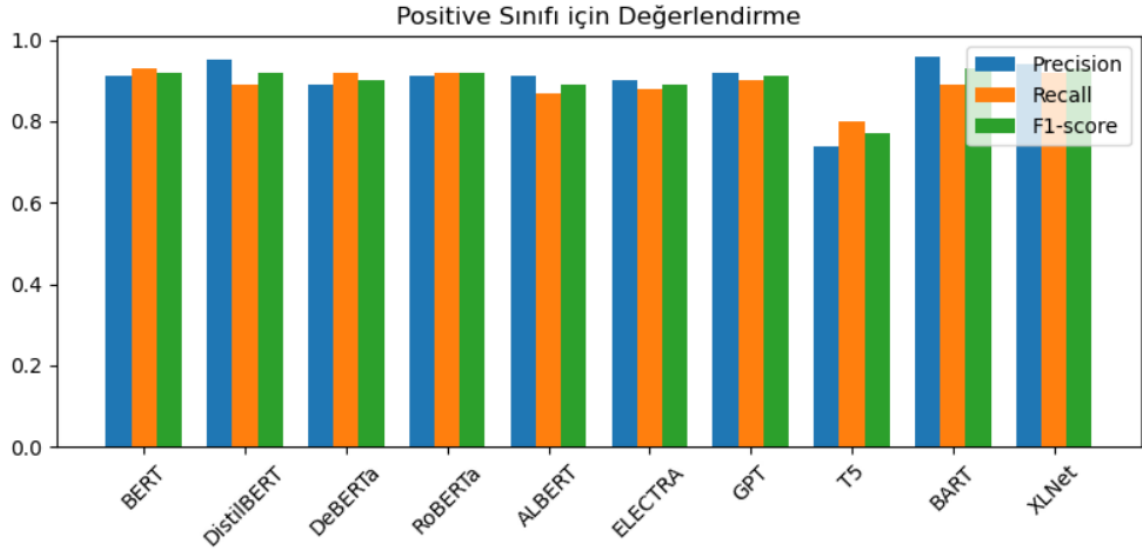
-Negatif sınıfta en yüksek precision değerini %95 oran ile BERT ve GPT modelleri almış ve en düşük değeri de %78 ile T5 modeli almıştır. Yine negatif sınıf tahmininde en yüksek recall değerini %96 ile DistilBERT modeli ve en düşük değeri %83 ile T5 modeli almıştır. Negatif sınıf için bir diğer değerlendirme metriği olan F1-

score için en yüksek değeri %93 oran ile BERT, RoBERTa ve GPT modelleri almış ve en düşük değeri ise %80 ile T5 değeri almıştır. Sonuçlar Şekil 4.9’da görselleştirilmiştir.



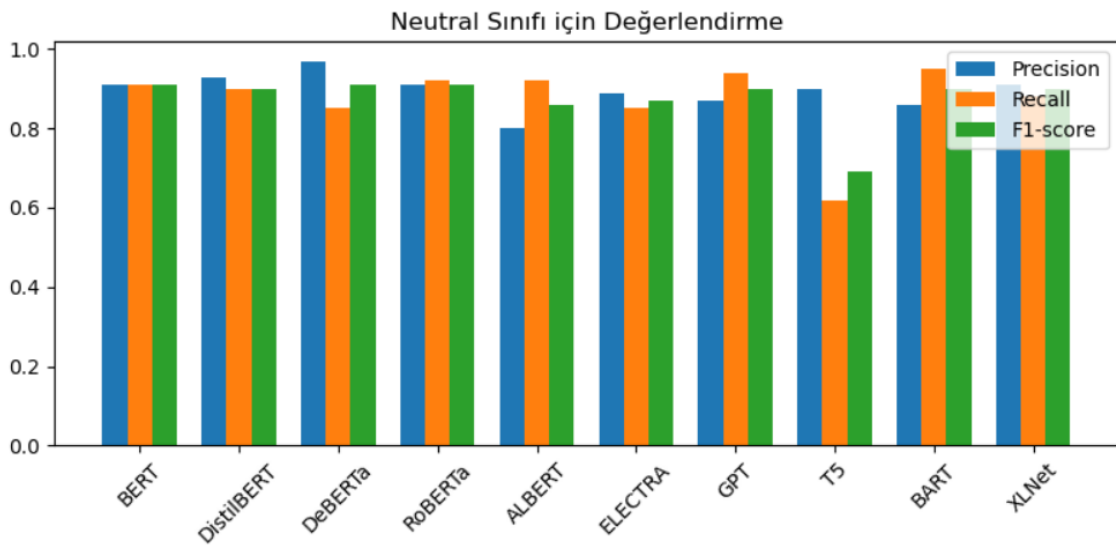
Şekil 4.9 Twitter (x) veri setinde negatif sınıfının değerlendirme metrikleri oranları

-Pozitif sınıf değerlendirmelerinde en yüksek precision değerini %95 oran ile DistilBERT modeli ve en düşük değeri %74 ile T5 modeli almıştır. En yüksek recall değerini %93 ile BERT modeli ve en düşük recall değerini %80 ile T5 modeli almıştır. F1-score için yapılan değerlendirmede ise en yüksek oranı %93 ile BART ve XLNet modelleri yakalamış ve %77 oranla T5 modeli en düşük değerde kalmıştır. Sonuçlar Şekil 4.10’da verilmiştir.



Şekil 4.10 Twitter (x) veri setinde pozitif sınıfının değerlendirme metrikleri oranları

-Nötr sınıfı karşılaştırmaları adına precision metriği için en yüksek başarıyı %97 ile DeBERTa modeli yakalamış ve %76 oran ile T5 modeli yine en düşük başarı oranında kalmıştır. Recall metriği için %95 ile BART modeli en yüksek oranı yakalamış ve %62 oranla T5 modeli en geride kalmıştır. Nötr sınıfının f1-score değerlendirmesinde ise BERT, DistilBERT, DeBERTa ve RoBERTa modelleri %91 oranla üst sırayı paylaşmış ve %69 oranla T5 modeli bütün modellerin gerisinde kalmıştır. Sonuçlar Şekil 4.11’de gösterilmiştir.



Şekil 4.11 Twitter (x) veri setinde nötr sınıfının değerlendirme metrikleri oranları

Twitter (x) veri seti için Transformer modellerinin sınıflar adına değerlendirme metrikleri Çizelge 4.21’de tablo halinde sunulmuştur.

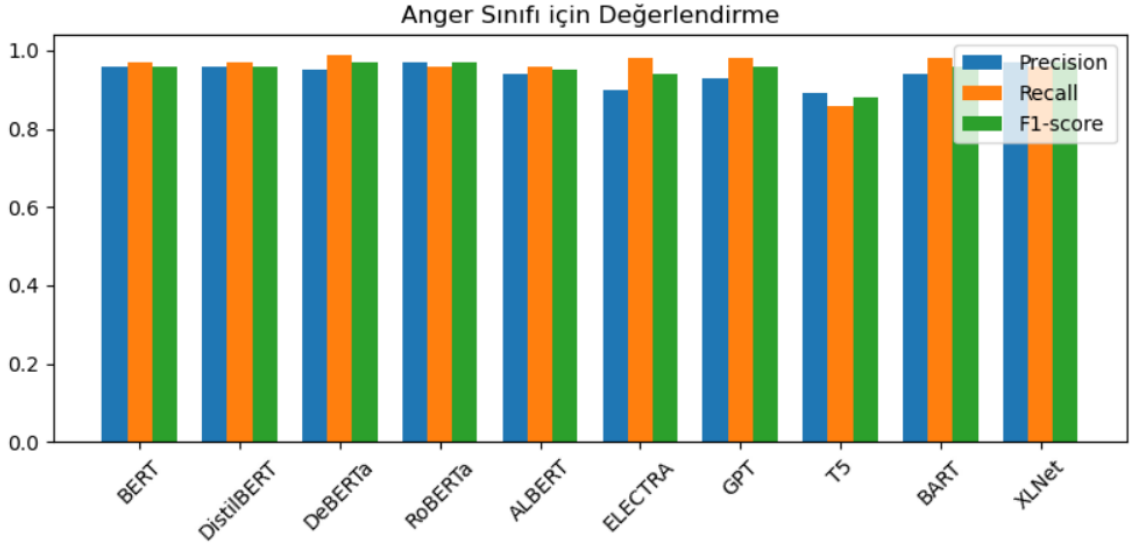
Çizelge 4.21 Twitter (x) veri setinde Transformer modellerinin negatif, pozitif ve nötr sınıflarında precision, recall ve f1-score oranı karşılaştırması

Model	Negative Precision	Negative Recall	Negative F1- score	Positive Precision	Positive Recall	Positive F1-score	Neutral Precision	Neutral Recall	Neutral F1-score
BERT	0.95	0.92	0.93	0.91	0.93	0.92	0.91	0.91	0.91
DistilBERT	0.89	0.96	0.92	0.95	0.89	0.92	0.93	0.90	0.90
DeBERTa	0.89	0.95	0.92	0.89	0.92	0.90	0.97	0.85	0.91
RoBERTa	0.93	0.92	0.93	0.91	0.92	0.92	0.91	0.92	0.91
ALBERT	0.94	0.86	0.90	0.91	0.87	0.89	0.80	0.92	0.86
ELECTRA	0.88	0.92	0.90	0.90	0.88	0.89	0.89	0.85	0.87
GPT	0.95	0.91	0.93	0.92	0.90	0.91	0.87	0.94	0.90
T5	0.78	0.83	0.80	0.74	0.80	0.77	0.90	0.62	0.69
BART	0.93	0.91	0.92	0.96	0.89	0.93	0.86	0.95	0.90
XLNet	0.88	0.93	0.90	0.94	0.92	0.93	0.91	0.88	0.90

4.2.2 Emotion Veri Seti

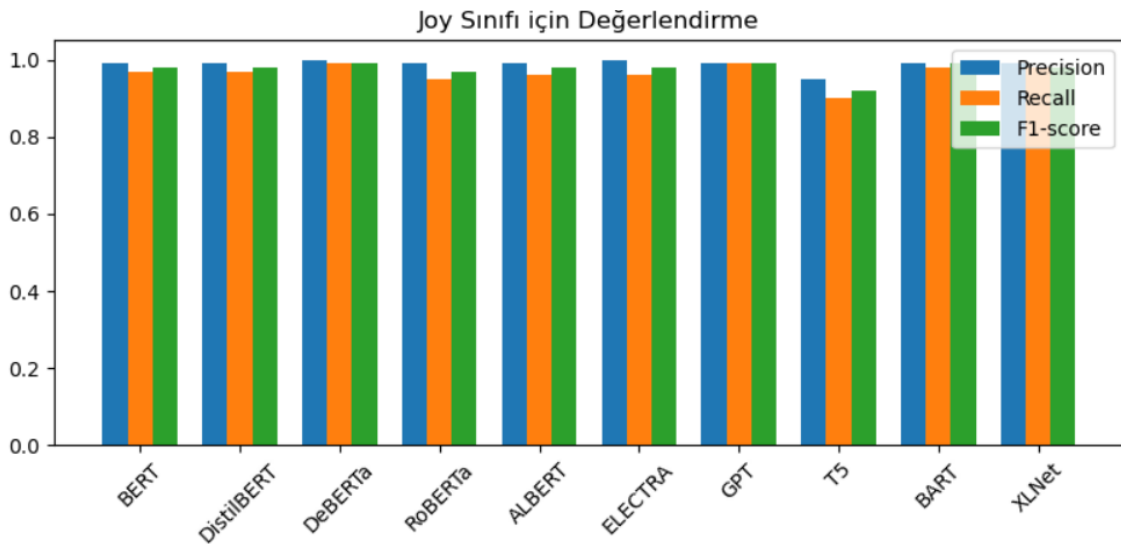
Emotion veri seti ile yapılan çalışmada ise metriklerle alakalı şu sonuçlara varılmıştır:

-Anger (kızgın) sınıfında en yüksek precision değerini %97 ile RoBERTa ve XLNet modelleri ve en düşük precision değerini ise %84 ile T5 değeri almıştır. Yine anger sınıfında recall değerini %99 oranla DeBERTa en yüksek, %86 ile de T5 modeli en düşük oranda temsil etmiştir. Anger sınıfının f1-score değerine baktığımızda %97 oranla DeBERTa, RoBERTa ve XLNet modelleri en yüksek değeri alırken %88 oran ile T5 modeli en düşük değeri vermiştir. Sonuçlar Şekil 4.12’de verilmiştir.



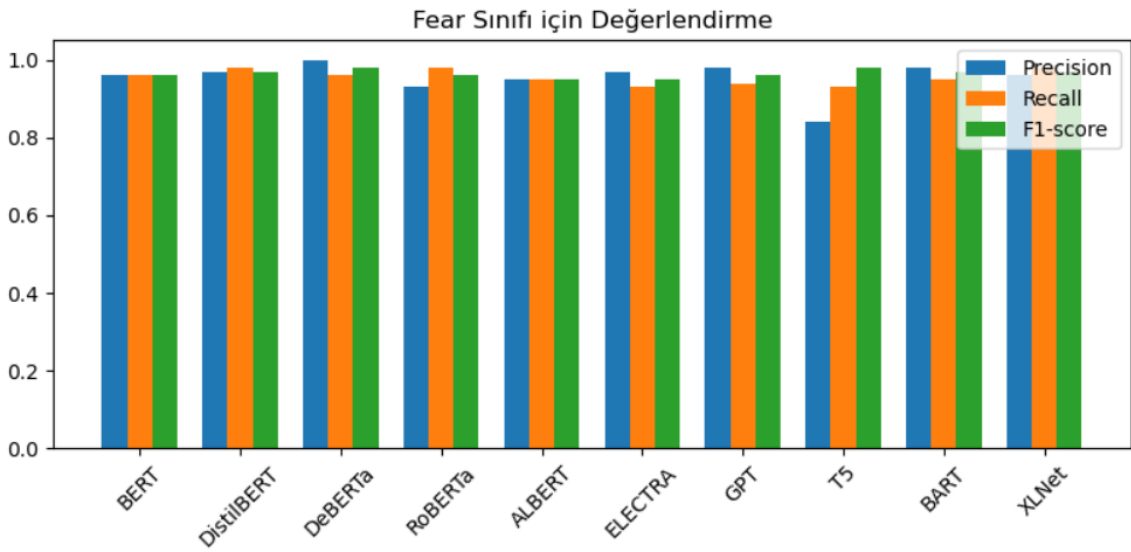
Şekil 4.12 Emotion veri setinde anger sınıfının değerlendirme metrikleri oranları

-Joy (neşeli) sınıfının değerleri için metriklere bakıldığında precision metriği için en yüksek oranı %100 ile DeBERTa ve ELECTRA modelleri, %95 ile en düşük oranı da T5 modeli almıştır. Joy sınıfı için recall değerlerine bakıldığında %99 oran ile DeBERTa ve GPT modelleri öne çıkmış ve %90 ile T5 modeli en düşük değerde kalmıştır. Joy sınıfının bir başka değerlendirme metriği olan f1-score için bakıldığında en yüksek değere %99 ile yine DeBERTa ve GPT modelleri ulaşmış ve en düşük oranda da %92 ile T5 modeli kalmıştır. Sonuçlar grafik halinde Şekil 4.13'te verilmiştir.



Şekil 4.13 Emotion veri setinde joy sınıfının değerlendirme metrikleri oranları

-Fear (korkmuş) etiketli değerler için precision en yüksek değeri %100 ile DeBERTa ve en düşük değeri %84 ile T5 olmuştur. Recall için bakıldığında %98 oranında DistilBERT, RoBERTa ve XLNet modellerinin en yüksek değere ulaştığı ve %93 oran ile ELECTRA ve T5 modellerinin en düşük değerleri aldığı görülmüştür. Fear sınıfının f1-score değeri için ise %98 ile DeBERTa öne çıkmış ve %88 ile T5 modeli en düşük değeri almıştır. Şekil 4.14'te bu sonuçlar görselleştirilmiştir.



Şekil 4.14 Emotion veri setinde fear sınıfının değerlendirme metrikleri oranları

Emotion veri seti için Transformer modellerinin sınıflar adına değerlendirme metrikleri Çizelge 4.22'de tablo halinde sunulmuştur.

Çizelge 4.22 Emotion veri setinde Transformer modellerinin anger, joy ve fear sınıflarında precision, recall ve f1-score oranı karşılaştırması

Model	Anger Precision	Anger Recall	Anger F1-score	Joy Precision	Joy Recall	Joy F1-score	Fear Precision	Fear Recall	Fear F1-score
BERT	0.96	0.97	0.96	0.99	0.97	0.98	0.96	0.96	0.96
DistilBERT	0.96	0.97	0.96	0.99	0.97	0.98	0.97	0.98	0.97
DeBERTa	0.95	0.99	0.97	1.00	0.99	0.99	1.00	0.96	0.98
RoBERTa	0.97	0.96	0.97	0.99	0.95	0.97	0.93	0.98	0.96
ALBERT	0.94	0.96	0.95	0.99	0.96	0.98	0.95	0.95	0.95
ELECTRA	0.90	0.98	0.94	1.00	0.96	0.98	0.97	0.93	0.95
GPT	0.93	0.98	0.96	0.99	0.99	0.99	0.98	0.94	0.96
T5	0.89	0.86	0.88	0.95	0.90	0.92	0.84	0.93	0.98
BART	0.94	0.98	0.96	0.99	0.98	0.99	0.98	0.95	0.97
XLNet	0.97	0.96	0.97	0.99	0.98	0.98	0.96	0.98	0.97

5. TARTIŞMA VE SONUÇ

Bu tez çalışması kapsamında yapılan literatür araştırması ve bulgular sonucunda Transformer tabanlı modeller ile ilgili çeşitli çalışmalar yapılmış fakat alanın çalışmaya çok açık olduğu da bütün çalışmalar sonucunda vurgulanmıştır. Alanın çalışmaya açık olması ve duygu analizinin dünyada ilgi alanı haline gelmesi buraya yoğunlaşma miktarını artırmış ve giderek artan oranda çalışmaların devam etmesini sağlamıştır. Gelişmeler alanın popüler kalmaya devam edeceğini ve buraya önemli yatırımların yapılmasının gerekli olduğunu gözler önüne sermiştir. Özellikle son dönemlerde daha fazla kullanılmaya başlanan Transformer tabanlı modeller bu alanda yenilikler ve yüksek başarı oranları getirmiş ve alanı daha popüler hale getirmiştir.

Duygu analizinde Transformer tabanlı modellerin kullanımı giderek artmakta ve bu modellere özgü özelleştirme çalışmaları da devam etmektedir. Bu çalışmada duygu analizi problemlerinde uygulanabilecek Transformer tabanlı modeller karşılaştırılmıştır. ALBERT, BART, BERT, DeBERTa, DistilBERT, ELECTRA, GPT, RoBERTa, T5 ve XLNet olmak üzere on ayrı Transformer modeli ile çalışma yapılmıştır. Biri 61121 örnekten oluşan Twitter (x) veri seti ve diğeri 5937 örneklemden oluşan emotion veri seti olmak üzere iki ayrı veri seti ile Transformer modelleri çalıştırılmıştır. Her bir notebook beş epoch değeri ile çalıştırılmıştır. Bu çalıştırmalar sonucunda her bir model için sonuçlar verilmiştir. Modellerle alakalı karmaşıklık matrisleri ve çeşitli grafikler de değerlendirilmiştir. Model performans sonuçları literatür taraması sonucunda bulunan çalışmalarla kıyaslandığında birçok modelin başarısının yükseldiği veya benzer başarılar yakaladıkları gözlemlenmiştir. Ayrıca Transformer tabanlı modellerin daha birçok çalışmaya açık olduğu ve daha çok geliştirilmesi gerektiği de belirlenmiştir. İnce ayar yapılandırmaları, modele özgü metrik kullanımı ve parametre dizaynı sonucunda gelecekte çok daha başarılı sonuçlar elde edilebileceği varsayılmaktadır. Çalışma sonunda genel olarak şu sonuçlar elde edilmiştir.

Modellerin büyük veri seti ve daha küçük bir veri seti ile alınan doğruluk karşılaştırmasında BERT modelinin büyük veri seti ile en yüksek doğruluk oranını verdiği ve küçük veri seti ile çalıştırıldığında diğer modellere göre başarısının daha az arttığı görülmüştür. Aynı şekilde büyük veri seti ile yapılan çalışmada T5 modeli en düşük başarıyı yakalarken küçük veri seti ile çalıştırıldığında diğer modellere kıyasla

başarısının daha fazla arttığı görülmüştür. Bu sonuçlar modellerin farklı büyüklükteki veri setlerinde verdikleri performans karşılaştırmasına güzel bir örnek olmuştur.

Twitter (x) veri seti ile yapılan çalışmada negatif sınıfının tahminlerine bakıldığında en yüksek precision değerini BERT ve GPT modellerinin, recall metriğinde en yüksek değeri DistilBERT modelinin ve en yüksek f1-score değerini ise BERT, RoBERTa ve GPT modellerinin aldığı ve üç metrik değerlendirmesinde de en düşük değeri T5 modelinin aldığı görülmüştür. Bu durumda T5 modelinin genel olarak düşük çıkmasının kendi model ayarlarından ve tokenizerinden kaynaklandığı ve sınıflandırma çalışmalarında daha genel ve dolaylı sonuçlar verdiği bir kez daha belirlenmiştir. Perikos ve Diamantopoulos, (2024) tarafından yapılan çalışmada yazarlar RoBERTa modelinin en yüksek f1-score değerine ulaştığını ve bu sonucun modelin çeşitli bağlamlarda duyguları doğru bir şekilde tanıma ve sınıflandırma yeteneğine kanıt olduğunu belirtmişlerdir. Yine aynı çalışmada Perikos ve Diamantopoulos, (2024) DistilBERT modelinin de olağanüstü performans göstermesinin dönüştürücü modellerin pratik faydalarını güçlendirdiğini ve büyük kaynak gereksinimi olmadan yüksek performans sergilediklerini belirtmişlerdir.

GPT modellerinin kavrayıcılığı, XLNet modellerinin yüksek performansı ve BERT ve türevleri olan çift yönlü çalışan modellerin gerek hız gerekse az kaynak kullanım durumları sayesinde büyük veri setleri ile hem daha az zamanda hem de yüksek başarı oranlarıyla gerçek dünya verilerine de uygulanabilirlik katsayısı artırılabilir. Perikos ve Diamantopoulos, (2024) yaptıkları çalışmada bu modellere ince ayar yapıp kullanıldıklarında yüksek oranda başarı yakalayabileceklerini ifade etmişlerdir. Böylece karmaşık gibi görünen problemler çözülebilecek ve insanların hayatını önemli ölçüde etkileyen ve onların hayatında büyük bir yer edinen sosyal medya, siyaset gibi mecralarda daha verimli zamanlar geçirilebilecektir. (Shetty vd., 2025) çalışmalarında özellikle XLNet modelinin başarısına değinmiş fakat eğitiminin yüksek süreler aldığını ve bunun da kaynak kısıtı olan durumlarda dezavantaj olabileceğini vurgulamışlardır. Bu çalışmada da XLNet modelinin özellikle Twitter (x) veri setinde eğitimi uzun süre almıştır. Ayrıca (Shetty vd., 2025) DistilBERT, ALBERT, RoBERTa, BERT, ELECTRA ve XLNet modelleri ile yaptıkları çalışmalarında en az eğitim süresine DistilBERT modeli ile ulaştıklarını belirtmişlerdir.

Transformer tabanlı modellerin duygu analizindeki performansları arasında gözlemlenen farklılıklar modellerin mimari yapıları, parametre büyüklükleri, ön eğitim stratejileri ve bağlamsal temsil yetenekleri gibi birçok etkene dayanmaktadır. Örneğin bu çalışmada BERT modeli Emotion veri setinde %96.88 gibi yüksek bir doğruluk oranı elde ederken ALBERT modeli daha düşük bir başarı sergilemiştir. Bu durum BERT'in çift yönlü bağlam temsilini güçlü bir şekilde gerçekleştirmesi ve tam boyutlu parametre setiyle yüksek öğrenme kapasitesi sunmasından kaynaklanmaktadır. Buna karşın ALBERT modeli parametre paylaşımı ve embedding boyutlarının azaltılması yoluyla daha hafif bir yapı sunmayı amaçlamaktadır. Bu sadeleştirilmiş yapı özellikle büyük ölçekli veri kümelerinde hesaplama verimliliği sağlarken küçük ve bağlamsal olarak zengin metinlerde detay kaybına yol açabilmektedir. Dolayısıyla ALBERT'in daha düşük başarı göstermesi mimarisinin yüksek doğruluk yerine kaynak verimliliği odaklı olmasından kaynaklanmaktadır. Benzer şekilde bazı modellerin düşük performans göstermesi yalnızca parametre sayısı değil, aynı zamanda dikkat (attention) mekanizmalarının yapılandırılma biçimi, eğitim süresince kullanılan optimizasyon teknikleri ve modele özel ön eğitim süreçleriyle de ilişkilidir. Özellikle DeBERTa ve RoBERTa gibi modeller içerik ve konum kodlaması gibi iyileştirilmiş mekanizmalar sayesinde belirli durumlarda daha güçlü bağlamsal kavrayış sergileyebilmiştir.

Twitter (x) veri seti için RoBERTa modelinin precision, f1-score gibi değerlerinin gayet başarılı olduğu ve Emotion veri seti için de precision, recall ve f1-score değerlerinin yüksek oranda başarılı olduğu gözlenmiştir. (El Azzouzy vd., 2025) yaptıkları çalışmada çalıştıkları tüm veri setleri üzerinde RoBERTa modelinin en iyi kesinlik, hatırlama ve F1 puanını aldığını belirtmişlerdir. Çalışmalarında özellikle kullanılan tokenizerler üzerinde durmuşlardır. Bu sonuçlar RoBERTa modelinin özel tokenizerleriyle kullanıldığında yüksek performans yakaladığı anlamına gelmektedir. (El Azzouzy vd., 2025) çalışmalarında ayrıca RoBERTa modelinin GPT, BERT ve T5 modellerine göre gelişmiş performansına vurgu yapmışlar ve bunun sebebinin RoBERTa modelinin daha büyük veri setleri üzerinde eğitilmesi ve maskelerin kaldırılması olduğunu savunmuşlardır. Ayrıca Twitter (x) veri setinde pozitif sınıflar için en yüksek precision değerini ve nötr sınıflarda da en iyi recall değerini yakalayan BART modeli aynı zamanda yüksek doğruluk oranlarıyla da başarılı bir performans yakalamıştır.

Bütün bu sonuçlar kapsamında Transformer modelleri arasındaki performans farklarının yalnızca doğruluk metrikleriyle değil modellerin mimari tasarımı ve veriyle etkileşim biçimiyle birlikte değerlendirilmeleri gerekmektedir. İleri düzey uygulamalarda model seçimi yapılırken yalnızca metrik performans değil donanım kaynakları, işlem süresi ve görev türüne uygunluk gibi çoklu kriterler göz önünde bulundurulmalıdır.

Bu çalışmada her bir Transformer modelinin farklı veri setleri üzerindeki performansı doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru gibi temel metrikler üzerinden değerlendirilmiştir. Ancak bu performans farklarının istatistiksel olarak anlamlı olup olmadığına yönelik bir hipotez testi gerçekleştirilmemiştir.

İnsan hayatı için vazgeçilemez hale gelen sosyal medya platformları, dijital içerik üreten mecralar ve alışverişin dijitalleşmesi gibi konular toplumların hayatına yerleşmiş hale gelmiştir. Böyle olunca buradaki içerikler doğru değerlendirilmeli ve ona göre tedbirler alınmalıdır. Gerek insan hakları konularında gerekse insanların diğer canlılarla olan etkileşimleri açısından bu son derece gerekli bir durumdur. Duygu analizi de bu anlamda gözden kaçırılmaması gereken konuların başında gelmektedir. Bu ortamlardaki içeriklerin duygu analizleri, içerdikleri hakaret, aşağılama vb. durumlar için mutlaka çalışmalar yapılması gereklidir. Bu konuda en iyi sonuçları verecek modellerin ve algoritmaların kullanımı ve bunların gerçek dünya verilerine uygulanması çalışmaları büyük bir hızla devam etmelidir. Böylece kötü sonuçların önüne geçilebilecek olup bu sonuçların en az kayıpla atlatılması konusunda başarılar elde edilebilecektir.

KAYNAKLAR

- Aljabar, A., Ali, I., Karomah, B. M. (2024). Sentiment analysis using transformer method. *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 6(2), 90-97. doi:10.20895/inista.v6i2.1383
- Almufareh, M. F., Jhanjhi, N. Z., Khan, N. A., Almuayqil, S. N., Humayun, M., Javed, D. (2024). BertSent: Transformer-based model for sentiment analysis of penta-class tweet classification. *IEEE Access*, 12, 196803-196817. doi:10.1109/ACCESS.2024.3515836
- Anonim, (2022). Python Nedir. Erişim tarihi: 16 Haziran 2025. Erişim adresi: <https://aws.amazon.com/tr/what-is/python/>
- Anonim, (2022). What is NLP. Erişim tarihi: 16 Haziran 2025. Erişim adresi: <https://amazon.com/insights/what-is-nlp-and-how-it-is-implemented-in-our-lives/>
- Anonim, (2025). NLP Nedir?. Erişim tarihi: 16 Haziran 2025. Erişim adresi: <https://tr.shaip.com/blog/what-is-nlp-how-it-works-benefits-challenges-examples/>
- Apu, Md. S. H., Islam, M. S., Aurpa, T. T. (2025). Explainable ai for sentiment analysis of human metapneumovirus (HMPV) using XLNet. *arXiv preprint arXiv:2502.01663*.
- Aydın, N., Erdem, O. A., Tekerek, A. (2024). Metin sınıflandırması için geleneksel makine öğrenimi ve dönüştürücü tabanlı derin öğrenme modellerinin karşılaştırmalı analizi. *Journal of Polytechnic*, 28(2), 445-452. doi:10.2339/politeknik.1469530
- Bandaranayake, I., Usoof, H. (2025). Sentiment Analysis of Sinhala News Comments Using Transformers. **Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages**, Abu Dhabi, UAE.
- Barak, B. (2021, 15 Aralık). *Duygu analizi nedir? nasıl yapılır?* Siberegitmen. <https://www.siberegitmen.com/duygu-analizi-nedir-nasil-yapilir/>
- Chifu, A. G., Fournier, S. (2023). Sentiment difficulty in aspect-based sentiment analysis. *Mathematics*, 11(4647), 1-33. doi:10.3390/math11224647
- Clark, K., Luong, M.-T., Le, Q. V., Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.
- Demir, B. D. (2019). Duygu analizi (sentiment analysis) nedir? Python ile uygulamalı duygu analizi ve fikir madenciliğine giriş (sentiment analysis & opinion mining). Erişim tarihi: 2 Şubat 2025. Erişim adresi: <https://ardabatuhandemir.medium.com/duygu-analizi-sentiment-anaylsis-nedir-68a59a8b0142>
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018a). BERT: Pre-training of deep bidirectional transformers for language understanding. **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)**, Minneapolis, Minnesota, USA.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018b). BERT: Pre-training of deep bidirectional transformers for language understanding. **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)**, Minneapolis, Minnesota, USA.
- Duan, G., Yan, S., Zhang, M. (2024). A Hybrid Neural Network Model for Sentiment Analysis of Financial Texts Using Topic Extraction, Pre-Trained Model, and

- Enhanced Attention Mechanism Methods. *IEEE Access*, 12, 98207-98224. doi:10.1109/ACCESS.2024.3429150
- El Azzouzy, O., Chanyour, T., Jai Andaloussi, S. (2025). Transformer-based models for sentiment analysis of youtube video comments. Available at SSRN 5071111.
- Eser, G., Sahin, C. (2024). Sentiment analysis and rating prediction for app reviews using transformer-based models. *International Journal of Advanced Natural Sciences and Engineering Researches*, 4(4), 372-379. <https://as-proceeding.com/index.php/ijanser>.
- Ethan, M. (2025). Comparative study of transformer-based models and Bi-LSTM for Bangla sentiment analysis using hybrid. <https://www.researchgate.net/publication/390877211>.
- Fadhil, A. (2024). Sentiment analysis using transformers. *Informatika dan Sains*, 14(3), 410-418.
- Fan, H., Du, W., Dahou, A., Ewees, A. A., Yousri, D., Elaziz, M. A., Elseikh, A. H., Abualigah, L., Al-Qaness, M. A. A. (2021). Social media toxicity classification using deep learning: Real-world application uk brexit. *Electronics (Switzerland)*, 10(1332), 1-18. doi:10.3390/electronics10111332
- Güven, Z. A. (2022). The comparison of language models with a novel text filtering approach for Turkish sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(2), 55. doi:10.1145/3557892
- Hamed, O., Zaidkilani, N. (2025). Exploring author style in nakba short stories: A comparative study of transformer-based models. Erişim tarihi: 5 Mart 2025. Erişim adresi: <https://www.nakba-archive.org>.
- He, P., Liu, X., Gao, J., Chen, W. (2020). DeBERTa: Decoding-enhanced bert with disentangled attention. arXiv 2020. *arXiv preprint arXiv:2006.03654*.
- Jahin, M. A., Shovon, M. S. H., Mridha, M. F., Islam, M. R., Watanobe, Y. (2024). A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *Scientific Reports*, 14(1), 24882. doi:10.1038/s41598-024-76079-5
- Mitri, A. M., Lyngdoh, E. L., Warjri, S., Saha, G., Lyngdoh, S. A., Maji, A. K. (2024). Probing a pretrained RoBERTa on Khasi language for POS tagging. *Natural Language Processing*, 31, 230-249. doi:10.1017/nlp.2024.24
- Orynbay, L., Bekmanova, G., Yergesh, B., Omarbekova, A., Sairanbekova, A., Sharipbay, A. (2024). The role of cognitive computing in NLP. *Frontiers in Computer Science*, 6, 1486581. doi:10.3389/fcomp.2024.1486581
- Özdamar, Y. (2020). Duygu analizi nasıl yapılır?-Doğal dil işleme. Teknoloji. Erişim tarihi: 15 Nisan 2025. Erişim adresi: <https://www.teknoloji.org/duygu-analizi-nasil-yapilir-dogal-dil-isleme/>
- Perikos, I., Diamantopoulos, A. (2024). Explainable aspect-based sentiment analysis using transformer models. *Big Data and Cognitive Computing*, 8(11), 1-30. doi:10.3390/bdcc8110141
- Riaz, A., Abdulkader, O., Ikram, M. J., Jan, S. (2025). Exploring topic modelling: A comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity. *International Journal of Electrical and Computer Engineering (IJECE)*, 15(2), 1933. doi:10.11591/ijece.v15i2.pp1933-1948

- Shaik Vadla, M. K., Suresh, M. A., Viswanathan, V. K. (2024). Enhancing product design through ai-driven sentiment analysis of Amazon reviews using BERT. *Algorithms*, 17(59), 1-17. doi:10.3390/a17020059
- Shetty, A. M., Manjaiah, D. H., Aljunid, M. F. (2025). Fine-tuning XLNet for Amazon review sentiment analysis: A comparative evaluation of transformer models. *ETRI Journal*. doi:10.4218/etrij.2024-0318
- Singh, U. (2024). Comparing transformer architectures for sentiment analysis: A study of BERT, GPT, and T5. *IJIRT 167124 International Journal of Innovative Research in Technology*, 11(3), 394-399.
- Suresh, S., Banu, K. S. (2024). Enhanced local and global context focus mechanism using bart model for aspect based sentiment analysis of social media data. *IEEE Access*. 12, 145768-145781.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. **31st Conference on Neural Information Processing Systems (NIPS 2017)**, Long Beach, CA, USA.
- Zhang, H., Shafiq, M. O. (2024). Survey of transformers and towards ensemble learning using transformers for natural language processing. *Journal of Big Data*, 11(1), 1-45. doi:10.1186/s40537-023-00842-0



ÖZ GEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Hayrullah TEMEL

Eğitim Bilgileri

Lisans

Üniversite : İnönü Üniversitesi

Fakülte : Mühendislik Fakültesi

Bölüm : Makine Mühendisliği

Mezuniyet Yılı : 2019





VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU

Tarih 16/06/2025

Tez Başlığı: Duygu Analizinde Transformer Tabanlı Modellerin Karşılaştırılması

Yukarıda başlığı belirtilen tez çalışmamın, kapak sayfası, giriş, ana bölümler ve sonuç bölümlerinden oluşan toplam 59 (elli dokuz) sayfalık kısmına ilişkin, 16/06/2025 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre tezimin benzerlik oranı %12 (on iki) dir.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre 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.

Gereğini bilgilerinize arz ederim.

Tarih ve İmza
16/06/2025

Adı Soyadı: Hayrullah TEMEL
Öğrenci No: 23910032010
Anabilim Dalı: Yapay Zeka ve Robotik
Programı: Yapay Zeka ve Robotik
Statüsü: (×) Yüksek lisans () Doktora

DANIŞMAN
UYGUNDUR

ENSTİTÜ ONAYI
UYGUNDUR