
GAZİ ÜNİVERSİTESİ
Gazi
1926

GENİŞ DİL MODELLERİNDE HİS ANALİZİ

Seda Nur ALTUN

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EYLÜL 2025

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Seda Nur ALTUN

15/09/2025

GENİŞ DİL MODELLERİNDE HİS ANALİZİ

(Yüksek Lisans Tezi)

Seda Nur ALTUN

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Eylül 2025

ÖZET

Doğal dil işleme alanındaki son gelişmeler, his analizi gibi dilin anlamını ve bağlamını derinlemesine gerektiren görevlerde önemli ilerlemeleri mümkün kılmıştır. Bu tez çalışması, geniş dil modellerinin bağlama dayalı his analizi görevlerindeki potansiyelini incelemekte ve farklı öğrenme senaryoları çerçevesinde değerlendirmeler sunmaktadır. Sıfır atışlı, az atışlı ve ince ayarlı yaklaşımlar temelinde yürütülen çalışmada, modelin bağlamsal duyarlılığı ve genelleme kapasitesi çeşitli istem türleri ve sınıf yapılarına göre sistematik biçimde analiz edilmiştir. Farklı duygu veri setleri üzerinden gerçekleştirilen deneysel çalışmalar, sınıf dengesizliği, örtük duyguların tanınması ve bağlam uzunluğunun etkisi gibi alanın temel zorluklarını ele almaktadır. Elde edilen bulgular, geniş dil modellerinin uygun yönlendirme ve özelleştirme stratejileriyle bu tür karmaşık görevlerde etkili bir biçimde kullanılabileceğini göstermektedir. Tez, his analizi alanında bağlamın belirleyici rolünü vurgulamakta ve gelecekteki çalışmalar için yol gösterici bir çerçeve sunmaktadır.

Bilim Kodu	:	92432
Anahtar Kelimeler	:	Geniş dil modelleri, his analizi, sıfır atış öğrenme, az atış öğrenme, ince ayar
Sayfa Adedi	:	81
Danışman	:	Doç. Dr. Murat DÖRTERLER

EMOTION ANALYSIS IN LARGE LANGUAGE MODELS

(M. Sc. Thesis)

Seda Nur ALTUN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

September 2025

ABSTRACT

Recent advances in natural language processing have enabled significant advances in tasks requiring in-depth understanding of language meaning and context, such as emotion analysis. This thesis examines the potential of large language models for context-based emotion analysis tasks and presents evaluations under various learning scenarios. Based on zero-shot, few-shot, and fine-tuned approaches, the model's contextual sensitivity and generalization capacity are systematically analyzed across various prompt types and class structures. Experimental studies conducted on diverse emotion datasets address key challenges in the field, such as class imbalance, implicit emotion recognition, and the effect of context length. The findings demonstrate that large language models can be effectively used in such complex tasks with appropriate manipulation and customization strategies. The thesis highlights the decisive role of context in emotion analysis and provides a guiding framework for future research.

Science Code	:	92432
Key Words	:	Large language models, emotion detection, zero shot learning, few shot learning, fine tuning
Page Number	:	81
Supervisor	:	Assoc. Prof. Dr. Murat DÖRTERLER

TEŞEKKÜR

Yüksek lisans eğitimim boyunca bana her türlü desteği sağlayan, vakit ayırıp her daim bana yol gösteren değerli hocam Doç. Dr. Murat DÖRTERLER'e teşekkürü bir borç bilirim. Eğitim ve öğretim hayatım boyunca her zaman yanımda olan, bu günlere gelmemde büyük emek veren kıymetli anne ve babama sevgi ve saygılarımı sunarım.

İÇİNDEKİLER

Sayfa

Sayfa

Sayfa

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil **Sayfa**

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
ÇKA	Çok Katmanlı Algılayıcı
DDİ	Doğal Dil İşleme
DRU	Düşük-Ranklı Uyumlama
ESA	Evrişimli Sinir Ağı
GDA	Grafik Dikkat Ağı
GDM	Geniş Dil Modeli
GESA	Graf Evrişimli Sinir Ağı
GİB	Grafik İşlem Birimi
GYB	Geçitli Yinelemeli Birim
PEİA	Parametre Etkin İnce Ayar
SDOKF	Sınıf Dengeli Odaklanmış Kayıp Fonksiyonu
TF-TBF	Terim Frekansı – Ters Belge Frekansı
UKSB	Uzun-Kısa Süreli Bellek
ÜRA	Üretici Rekabetçi Ağlar
YSA	Yinelemeli Sinir Ağı

1. GİRİŞ

His analizi metin, video, ses, görüntü gibi çoklu veri türleri üzerinde ifade edilen duygusal durumları otomatik olarak tespit edilmesi ve ardından ifade yoğunluğunun belirlenmesi olarak tanımlanır [1]. Gelişen teknoloji ile sağlık, bankacılık, eğitim gibi alanlarda insan bilgisayar etkileşiminin artması, his analizini daha önemli hale getirmiştir. Bunun sebebi kullanıcılar üzerinde daha etkili bir izlenim oluşturmak için yapay çıktılar yerine duygusal olarak da kullanıcıyı tatmin eden bir iletişim oluşturma gereksinimidir [2]. Oluşan bu talebi karşılayabilmek için ses, video, metin gibi çeşitli veri formatları üzerinde his analizi çalışmaları yapılmaktadır. Bu çalışmalar neticesinde kullanıcıların duygu durumlarını daha iyi anlayarak anlamlı iletişim sağlayan bilgisayar tabanlı çözümlerin hisleri tespit etmesi ve tespit ettiği duygu bağlamında yanıt vermesi hedeflenmektedir. Bunun yanı sıra his analizi uygulamaları kullanıcı memnuniyetini artırmak ve iletişim kalitesini iyileştirmek amacıyla sosyal medya, müşteri geri bildirimleri, ürün incelemeleri ve çağrı merkezi kayıtları gibi büyük miktarlarda verinin analizinde de kullanılmaktadır.

Duygu analizi ve his analizi terimleri benzer anlamlar taşımaları sebebiyle birbiriyle karıştırılabilmektedir. Duygu analizi pozitif, negatif veya nötr olarak sınıflandırma sürecini ifade eder. Gelişimin ilk aşamalarında duygu analizi pozitif ya da negatif olmak üzere daha temel seviyede yapılırken ilerleyen çalışmalarda çok pozitif, pozitif, nötr, negatif ve çok negatif gibi daha detaylı sınıflandırmalar yapılmıştır [3, 4]. His analizi ise duyguları daha ayrıntılı bir şekilde neşe, korku, öfke, üzüntü, tiksinme, şaşırma gibi sınıflara ayırarak tespit etmeyi amaçlamaktadır [5].

His analizi yöntemleri geleneksel yöntemler ve modern yöntemler olarak iki ana başlık altında ele alınabilir. Geleneksel yaklaşımlar; sözlük tabanlı yaklaşımlar, kural tabanlı yaklaşımlar, istatiksel yaklaşımlara dayanmaktadır. Bu yöntemler sınırlı kelime dağarcığı ve kelimelerin içinde bulundukları bağlamsal yapıyı dikkate almamaları nedeniyle karmaşık duygusal ifadeleri tespit etmede yetersiz kalmaktadırlar [6]. Bununla birlikte, artan veri hacmi karşısında verimli işlem yapamama ve ölçeklenebilirlik eksikliği, bu yaklaşımların büyük veri setleriyle çalışırken önemli bir dezavantaj oluşturmasına yol açmaktadır. Bu sınırlılıkların üstesinden gelmek amacıyla modern yöntemler başlığı altında yer alan makine öğrenimi ve derin öğrenme temelli yöntemler his analizinde yaygın

olarak kullanılmaya başlanmıştır. Derin öğrenme tabanlı modeller, bağlamı da dikkate alarak ironi içeren veya dolaylı ifadeleri de daha isabetli olarak sınıflandırabilme potansiyeli sunmaktadır. Özellikle dönüştürücü mimarisi ile birlikte geliştirilen modeller, metinler arasındaki uzun menzilli bağımlılıkları etkili şekilde modelleyebildiğinden bağlamı daha iyi bir şekilde kavrayabilmektedir.

His analizindeki bu gelişmelerin devamı olarak ortaya çıkan geniş dil modelleri (GDM) ise önemli performans artışları göstermiştir. Özellikle, dönüştürücü tabanlı modeller olan BERT [7], GPT [8] ve RoBERTa [9], metni anlama ve işleme konusundaki üstün yetenekleri sayesinde his analizde yeni bir dönem başlatmışlardır. Bu modeller, büyük veri kümeleri üzerinde önceden eğitilerek kapsamlı bir dil bilgisi edinmekte ve ardından his analizi gibi özel görevler için ince ayar yapılarak yüksek doğruluk oranları elde edebilmektedirler [10]. Üstelik bu modeller sıfır atışlı, az atışlı ve ince ayar gibi farklı öğrenme stratejilerini de destekleyerek kısıtlı veri senaryolarında da esnek çözümler sunmaktadırlar.

Literatür incelendiğinde his analizi alanında çok sayıda çalışma bulunmakla birlikte GDM'lerle sistematik performans analizi yapılmış çalışmaların sınırlı olduğu görülmektedir. Bu tez çalışmasında GDM'lerin his analizi görevindeki etkilerini incelemek için 3 farklı veri seti üzerinde bağlama dayalı his analizi yapılmıştır. GDM olarak Meta tarafından geliştirilen LLaMA 3 8B Instruct modeli kullanılarak sıfır atış, az atış ve ince ayar yaklaşımlarının his analizi performansı karşılaştırmalı olarak incelenmiştir.

2. LİTERATÜR ARAŞTIRMASI

Doğal dil işleme (DDİ) alanında metin tabanlı his analizi, önemli bir araştırma konusu olarak yer almaktadır. Bu alandaki çalışmalar başlangıçta anahtar kelime ve kural tabanlı yöntemler kullanılarak yapılmıştır ancak bu yöntemler metinde verilen bağlamı dikkate almakta yetersiz kaldıkları için karmaşık duygusal ifadeleri analiz etmekte yeterli olamamışlardır. Bu nedenle, makine öğrenmesi ve derin öğrenme tabanlı yöntemler ön plana çıkarak tercih edilmeye başlanmış ve his analizi görevlerinde başarıyı önemli ölçüde artırmıştır. Özellikle denetimli öğrenme ile eğitilen modeller, metinlerdeki bağlamsal yapıyı daha doğru bir şekilde yorumlayarak anlamlı gelişmeler sunmuştur.

Derin öğrenme modelleri içerisinde yer alan evrişimli sinir ağları (ESA) girdi verisini filtreleyen evrişim ve alt örnekleme katmanlarından oluşan ileri beslemeli bir yapay sinir ağı yapısıdır. Bu mimari metin sınıflandırma görevlerinde yüksek başarı oranlarıyla dikkat çekmektedir [11]. Li ve arkadaşları [12], Çin sosyal medya platformu Weibo üzerinde oluşturulan EACWT veri kümesi üzerinde Plutchik his modeli ile ESA tabanlı bir model geliştirerek kelime düzeyinde %42,8 cümle düzeyinde ise %23,6 F1 skoru elde etmişlerdir. Toçoğlu ve arkadaşları [13], Türkçe tweet verisi kullanarak Ekman his modelini kullanarak yapay sinir ağı, ESA ve uzun-kısa süreli bellek (UKSB) modellerini karşılaştırmış ve sırasıyla %71,42; %74,0 ve %72,28 doğruluk oranlarına ulaşmışlardır. Park ve arkadaşları [14] metinde ifade edilen hissin tanınması için yaptıkları çalışmalarında Plutchik his modelini kullanarak Twitter ve ROCStories veri kümeleri üzerinde ESA ile %51,25 doğruluk oranı elde etmişlerdir. Dheeraj ve arkadaşları [15] ruh sağlığı ile ilgili metinlerdeki olumsuz hisleri tespit etmek amacıyla MHA-BCNN isimli bir derin öğrenme modeli oluşturmuşlar ve WebMD ile HealthTap çevrimiçi sitelerinden elde ettikleri hasta metinlerinden oluşan veri setleri üzerinde sırasıyla %78,3 ve %89,5 makro F1 skoru elde etmişlerdir.

His analizi çalışmalarında kullanılan bir diğer model de yinelemeli sinir ağlarıdır (YSA). Bu yapay sinir ağı modeli sıralı veriler üzerinde çalışmak üzere tasarlanmıştır ve veri dizisindeki önceki öğelere dair bilgiyi de modelin belleğinde tutarak bağımlılıkları yakalayabilme yeteneğine sahiptir. Her bir girdiyi işlerken önceki adımlardan gelen bilgiyi de dikkate alarak bağlamsal ilişkilerin daha derin düzeyde öğrenilmesini sağladığından his

analizi gibi ardışık yapıdaki DDİ görevlerinde yaygın şekilde kullanılmaktadır. Modelin temel yapısında sinir hücreleri arasında döngüsel bağlantılar yer almakta ve bu geri besleme mekanizması sayesinde geçmiş girdilerden elde edilen bilgiler sonraki adımlardaki çıktıları etkilemektedir [16]. Bu çerçevede Yang ve arkadaşları [17] çalışmalarında Plutchik'in his modelini kullanarak YSA tabanlı ve dikkat mekanizması içeren bir model geliştirmişler ve Sina Social News, Ren-CECps, SemEval veri setleri üzerinde sırasıyla %73,79; %63,04; %75,38 mikro F1 skorları elde etmişlerdir. Majumder ve arkadaşları [18] ise DialogueRNN modelini oluşturarak IEMOCAP veri kümesi üzerinde çift yönlü YSA ve dikkat mekanizmasını birleştirerek oluşturdukları model ile %62,75 F1 skoru elde etmişlerdir.

UKSB ise sıralı verilerle çalışmak üzere tasarlanmış YSA mimarisinin geliştirilmiş halidir. YSA'larda uzun dizilerde önceki bilgilerin zamanlar silinmesi sorununu çözmek üzere tasarlanan UKSB geçmiş bilgileri hem kısa hem de uzun süreli bellekte tutabilme kapasitesine sahiptir. Giriş, unutma ve çıkış olmak üzere üç temel kapı mekanizmasına sahip olmaları sayesinde hangi bilginin saklanıp hangisinin unutulacağına dinamik biçimde karar verebildiğinden bağlama dayalı his analizinde daha doğru tahminler yapılmasına olanak sağlamaktadır [19]. Li ve arkadaşları [20] Ekman modelini temel alan DailyDialog ve Semeval19Emocn veri kümeleri üzerinde yaptıkları çalışmada çift yönlü UKSB ve dikkat katmanı kullanarak sırasıyla %38,85 ve %36,23 F1 skoru elde etmişlerdir. Kadan ve arkadaşları [21] yine Ekman his modelini kullanarak REN-20k, RENh-4k ve SemEval-2007 veri kümeleri üzerinde Bi-LSTM ve dikkat mekanizması ile %76,68; %60,75 ve %66,96 doğruluk oranları elde etmişlerdir. Batbaatar ve arkadaşları [22], DailyDialog, CrowdFlower, TEC ve ISEAR gibi çeşitli veri kümeleri üzerinde ESA ve çift yönlü UKSB modellerini test etmiş ve DailyDialog veri setinde %84,8 doğruluk elde etmişlerdir. Hu ve arkadaşları [23], UKSB ve dikkat mekanizmasını birleştiren DialogueCRN modelini IEMOCAP ve MELD veri kümelerinde test etmiş ve sırasıyla %66,20 ve %60,73 ağırlıklı F1 skorlarına ulaşmıştır. Aynı araştırma grubu daha sonra SACL-UKSB modelini geliştirerek IEMOCAP, MELD ve EmoryNLP veri kümelerinde sırasıyla %69,22; %66,45 ve %39,65 ağırlıklı F1 skorları elde etmiştir [24]. Liu et al. [25] tarafından geliştirilen EmotinIC modeli, duyguların sürekliliğini ve bilişsel özellikleri dikkate alarak bağlamsal analiz yeteneğini genişletmiş; IEMOCAP, MELD ve EmoryNLP veri kümelerinde sırasıyla %69,61, %66,32 ve %40,25 ağırlıklı F1 skorlarına ulaşmıştır. Ayrıca DailyDialog veri setinde %54,19 makro F1 skoru elde edilmiştir.

Metin tabanlı his analizinde kullanılan bir diğer derin öğrenme modeli ise geçitli yinelemeli birimlerdir (GYB). YSA yapısının daha sadeleştirilmiş ve verimli bir çeşidi olan GYB özellikle uzun süreli bağımlılıkların modellenmesinde klasik YSA'lara kıyasla daha başarılı sonuçlar vermektedir ve yapısı gereği daha az hesaplama maliyetiyle uzun bağlam ilişkilerinin modellenebilmesine olanak tanımaktadır [26]. Ghosal ve arkadaşları [27] IEMOCAP veri seti üzerinde GYB ile birlikte graf evrişimli sinir ağı (GESA) kullanarak %64,18'lik bir F1 skoru elde etmişlerdir. Lu ve arkadaşları [28] ise Ekman his modelini kullandıkları çalışmalarında MELD ve IEMOCAP veri setleri üzerinde çift yönlü GYB ile %60,72 ve %64,37 F1 puanı elde etmişlerdir. Ghosal ve arkadaşları [29] genel bilginin kullanımı yoluyla duygu tanımayı geliştirmeyi amaçladıkları çalışmalarında Plutchik modelini kullanarak DailyDialog, MELD, EmoryNLP ve IEMOCAP veri kümelerinde çift yönlü GYB ile %51,05; %65,21; %38,11 ve %65,28 F1 skorları elde etmiştir. Bir diğer çalışmada ise Jiao ve arkadaşları [30] tarafından geliştirilen HiGRU modeli Friends, EmotionPush ve IEMOCAP veri setlerinde sırasıyla %74,4; %77,1 ve %82,1 F1 skorları elde etmiştir. Mohammadi ve arkadaşları [31] yaptıkları çalışmada GYB tabanlı hibrit bir yaklaşım önermişler ve EmoContext veri setinde %69,93 F1 skor elde etmişlerdir. Akhtar ve arkadaşları [32] Plutchik modeline uygun olarak duyguları ve yoğunluklarını tahmin etmek için UKSB, ESA ve GYB kullanarak EmoInt-2017 veri setinde %78,8 doğruluk oranı elde etmişlerdir.

Graf Dikkat Ağları (GDA) ise metindeki kelimelerin bağlamsal ilişkilerini dikkate alarak her kelimenin anlamını durumuna göre esnek biçimde güncelleyebilen bir dikkat mekanizması içermektedir. Ishiwatari ve arkadaşları [33] ise yine Plutchik his modelini kullanarak DailyDialog, MELD, EmoryNLP ve IEMOCAP veri kümeleri üzerinde ilişkisel farkındalığa sahip GDA ile %54,31; %60,91; %34,42 ve %65,22 F1 skorları elde etmişlerdir. Shen ve arkadaşları [34], konuşmalarda his tanıma için yönlendirilmiş çevrimsiz grafik tabanlı model kullanarak IEMOCAP, MELD, DailyDialog ve EmoryNLP veri setlerinde sırasıyla %68,03; %63,65; %59,33 ve %39,02 F1 skorları raporlamıştır.

Metin tabanlı his analizinde yaygın olarak kullanılan bir diğer yöntem ise dönüştürücü tabanlı modellerdir. Bu modeller kodlayıcı ve çözücü olmak üzere iki ana bileşenden oluşur ve çıktı olasılıklarını düzeltmek için softmax fonksiyonu kullanırlar. BERT, XLNet, RoBERTa, SensoryT5, DistilBERT metin tabanlı his analizinde tercih edilen dönüştürücü tabanlı modellerdir. Zhong ve arkadaşları [35] çok başlıklı dikkat mekanizmasıyla birlikte

dönüştürücü tabanlı mimariyi birleştirerek EC, DailyDialog, MELD, EmoryNLP ve IEMOCAP veri kümeleri üzerinde deneyler gerçekleştirmiş ve EC veri setinde %73,4 ile en yüksek F1 skorunu elde etmişlerdir. Calò ve arkadaşları [36] Plutchik his modelini kullanarak İtalyanca metinlerdeki hisleri ayırt etmek için BERT modelini kullanmış ve %86,32'lik bir F1 skoru elde etmişlerdir. Huang ve arkadaşları [37] konuşmaya dayalı his analizi için EmotionBERT modelini önermişler ve bu model ile EmotionLines ve EmotionPush veri kümelerinde sırasıyla %81,5 ve %88,5 F1 skorları elde etmişlerdir. Li ve arkadaşları [38] konuşmadaki bağlam ve konuşmacıya duyarlılık konularını ele alarak MELD, EmoryNLP, ve IEMOCAP veri kümeleri üzerinde his analizi yapmışlar ve sırasıyla %61,94; %36,75 ve %64,5 F1 skorları elde etmişlerdir. Li ve arkadaşları [39], hiyerarşik dönüştürücü tabanlı mimari ve BERT mimarilerini kullandıkları ve Ekman modelini temel aldıkları çalışmalarında Friends, EmoryPush ve EmoryNLP veri kümeleri üzerinde deneylerini yapmışlar ve sırasıyla %67,88; %65,43 ve %33,04 F1 skoru elde etmişlerdir. Nugroho ve Bachtiar [40] Ekman duygu modeli temel alınarak Endonezce tweet'ler üzerinde gerçekleştirdikleri çalışmada BERT modeliyle %76,73 doğruluk oranı elde etmişlerdir. Huang ve arkadaşları [41], bağlamsal his tespiti alanında hiyerarşik yöntemlerin önemini vurguladıkları çalışmalarında EmoContext veri kümesi üzerinde UKSB ve BERT modellerinin birleşimini kullanarak %77,09 F1 skoru elde etmişlerdir. Al-Omari ve arkadaşları [42] ise, Ekman modelini temel alarak EmoContext veri kümesinde BERT ve çift yönlü UKSB kombinasyonu ile %74,78 F1 skoru elde etmişlerdir. Khurdula ve arkadaşları [43] ise, çift yönlü UKSB, BERT, dikkat mekanizması, Palm V2 ve Gemini Pro modellerinden oluşan bir sistemle, Twitter API'si üzerinden toplanan veri setinde sırasıyla %96 ve %92 doğruluk oranlarına ulaşmışlardır. Ancak Palm V2 ve Gemini Pro modellerine ait spesifik sonuçlar rapor edilmemiştir. Bu çalışma, çoklu derin öğrenme modellerinin birleşimiyle metin tabanlı his analizinin değerlendirilmesini amaçlamaktadır. Widarmanti ve arkadaşları [44] youtube yorumlarındaki hissi tespit etmek amacıyla yaptıkları çalışmalarında RoBERTa modelini kullanarak %64 doğruluk oranı elde etmişlerdir. Acheampong ve arkadaşları [45] farklı modellerin metin tabanlı his analizi üzerindeki başarımlarını karşılaştırmak istemişler ve Ekman his modeli ile ISEAR veri kümesi üzerinde BERT, RoBERTa, DistilBERT ve XLNet modellerini test ederek sırasıyla %70,09; %74,31; %66,93 ve %72,99 doğruluk oranları elde etmişlerdir. Zhao ve arkadaşları [46] ise çalışmalarında GoEmotions, EmD, ISEAR ve EmoInt veri kümelerinde SensoryT5 ile sırasıyla %67,0; %61,5; %72,4 ve %87,6 F1 skorlarını elde etmişlerdir. Song ve arkadaşları [47] tarafından geliştirilen SPCL+CL modeli, BERT tabanlı yapısıyla

%69,74; %67,25 ve %40,0 F1 skorlarına ulaşmıştır. Zhong ve arkadaşları [48] tarafından önerilen KET modeli, dış bilgi kaynaklarını da kullanarak MELD, EmoryNLP ve IEMOCAP veri kümelerinde sırasıyla %58,18; %34,39 ve %59,56 ağırlıklı F1 skorları elde etmiştir. Aynı model EC ve DailyDialog veri setlerinde sırasıyla %73,48 ve %53,37 mikro F1 skorları raporlamıştır. Li ve arkadaşları [49] tarafından geliştirilen CoG-BART ve Zhu ve arkadaşları [50] tarafından sunulan TODKAT modelleri ise üretici yaklaşımlar ile duygu sürekliliğini sağlama amacı taşımaktadır. CoG-BART modeli DailyDialog, MELD, IEMOCAP ve EmoryNLP veri kümelerinde sırasıyla %56,09; %64,81; %66,18 ve %39,04 ağırlıklı F1 skorları; TODKAT modeli ise sırasıyla %52,56; %68,23; %61,33 ve %43,12 F1 skorları raporlamıştır. Gou ve arkadaşları [51] ise his analizi görevi için konu, his ve dış bilgi temsillerini birleştiren TG-ERC DailyDialog, MELD, IEMOCAP, and EmoryNLP veri setleri üzerinde deneylerini gerçekleştirmişler ve sırasıyla %52,56; %70,81; %65,41 ve %42,38 F1 skorları elde etmişlerdir.

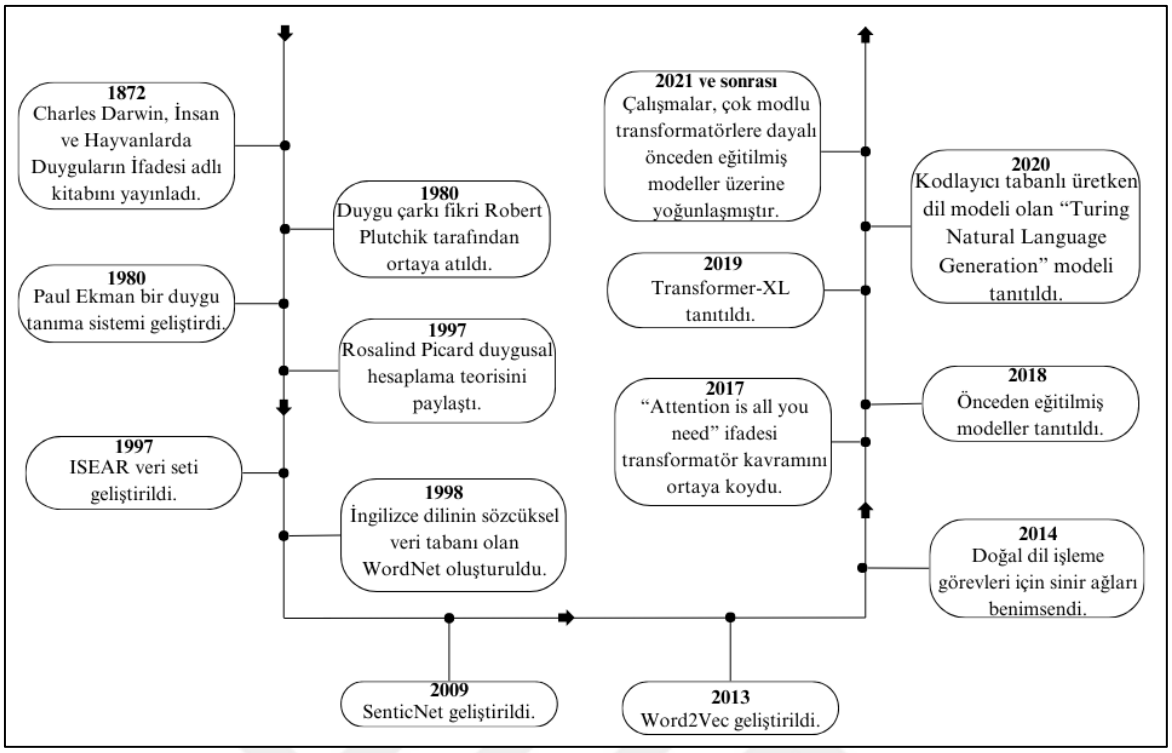
Bu tez çalışmasında ise, GDM’lerin metinlerdeki his tespitindeki başarısı sistematik olarak değerlendirilmiştir. Çalışmanın temel amacı, GDM’lerin ince ayar, az atışlı öğrenme, sıfır atışlı öğrenme gibi farklı öğrenme yaklaşımları altında sergiledikleri performansları karşılaştırarak, hangi yöntemin hangi koşullar altında daha etkili sonuçlar verdiğini ortaya koymaktır. Literatürde, çoğu çalışma yalnızca tek bir öğrenme yöntemine odaklanmakta veya farklı yaklaşımlar arasındaki karşılaştırmaları sınırlı düzeyde sunmaktadır. Oysa günümüzde GDM'ler, DDİ alanında çığır açan başarılar elde etmiş ve birçok uygulama alanında üstün performans sergilemiştir. Ancak bu modellerin his analizi gibi daha spesifik görevlerdeki etkisi, yeterince derinlikli biçimde incelenmemiştir. Bu tez çalışması ise söz konusu boşluğu doldurmayı amaçlamakta ve GDM’lerin his analizi bağlamındaki potansiyelini kapsamlı bir şekilde incelemeyi hedeflemektedir.

3. KURAMSAL VE KAVRAMSAL ÇERÇEVE

Bu bölümde, his analizine yönelik kuramsal temeller, alandaki tarihsel gelişim çizgisi ile birlikte sistematik biçimde ele alınmaktadır. İlk olarak hislerin tanımlanması ve modellenmesine ilişkin kuramsal yaklaşımlar aktarılmış, ardından bu alanda yaygın olarak kullanılan veri setleri ve veri işleme süreçleri detaylandırılmıştır. Ayrıca, metin tabanlı his analizinde uygulanan özellik çıkarımı teknikleri ile yöntemsel yaklaşımlar hem geleneksel hem de modern perspektiflerle değerlendirilmiştir. Bu kapsamda sunulan kuramsal çerçeve, tez çalışmasının deneysel bileşenlerine teorik zemin oluşturmayı amaçlamaktadır.

3.1. His Analizinin Evrimi

His terimine ilk olarak Charles Darwin tarafından yazılan “The Expression of Emotions in Man and Animals” adlı eserde yer verilmiştir [52]. Bu eser insan duygu ve ifadelerinin inceliklerini tanımlayan ilk teori olarak kabul edilmektedir. Bu alandaki en önemli gelişmelerden bir diğeri ise Plutchik tarafından geliştirilen his çarkıdır [53]. Plutchik sekiz temel hissi tanımlamış ve ayrıca her temel hissin ters bir karşılığı olduğunu önermiştir. Daha sonra ise Paul Ekman tarafından yüz ifadelerine dayanan his tanıma sistemi geliştirilmiştir [54]. Bu gelişmelerin ardından his analizi alanında yapılan araştırmalar yüz ifadelerinin ötesine geçerek konuşma temelli analizlere doğru genişlemiştir. His analizinin zaman içindeki gelişimi Şekil 3.1’de genel hatlarıyla gösterilmektedir.



Şekil 3.1. His analizinin gelişimi

1997 yılında Rosalind Picard'ın duygusal hesaplama teorisinin yayınlanması bu alanda önemli bir sıçramaya sebep olmuştur [55]. Picard'a göre bilgisayarların hisleri tanıma, yorumlama ve ifade etme kapasitesi insanlarla doğal bir iletişim halinde olabilmeleri için oldukça önemlidir. Ortaya atılan bu fikir duygusal hesaplama alanındaki araştırmaların artışına zemin hazırlamıştır. 1997'de ilk his veri seti olan ISEAR, 1998'de WordNet [56] DDİ görevleri için öncü bir sözcük veritabanı olarak ortaya çıkmıştır. 2009 yılında geliştirilen SenticNet [57] kavramsal düzeyde his analizine olanak sağlayan kamuya açık bir kaynak olarak sunulmuştur. Bu dönemde kelime gömme yöntemleri de önem kazanmaya başlamıştır. Özellikle 2013 yılında geliştirilen Word2Vec modeli [58], kelimeleri çok boyutlu vektörler olarak temsil ederek duygusal anlamların daha doğru biçimde yakalanmasına olanak sağlamıştır. 2014 yılında, sinir ağlarını ilk kez DDİ görevlerine uygulayarak his analizinde derin öğrenme tabanlı yeni bir dönemin başlangıcını yapmışlardır [59, 60]. 2017 yılında önerilen dönüştürücü tabanlı mimari [61], bağlamı çok daha etkili şekilde kavrama yeteneği sayesinde his analizi de dahil olmak üzere birçok DDİ görevinde çığır açmıştır [7]. 2018'den itibaren önceden eğitilmiş modellerin his analizinde kullanımı yaygınlaşmış, metin sınıflandırma ve analizde doğruluk oranlarında ciddi artışlar elde edilmiştir. 2019 yılında Google tarafından, bağlamın daha geniş kesitlerini anlayabilmek amacıyla Transformer-XL modeli

tanıtılmıştır [62]. 2020'de ise Microsoft, 17 milyar parametre içeren Turing Natural Language Generation adlı modelini tanıtarak şimdiye dek yayımlanmış en büyük modellerden birini ortaya koymuştur [63]. 2021 yılı itibarıyla araştırmalar, büyük önceden eğitilmiş modellere çok modlu veri entegre edilmesine odaklanmıştır. Bu gelişmeler, yalnızca metne dayalı his analizinin ötesine geçerek ses, görsel ve diğer veri türlerini de içeren sistemlerin doğmasına olanak sağlamış ve his analizinin daha geniş alanlarda uygulanabilir hale gelmesine öncülük etmiştir [64]. Karmaşık his analizi sistemleri artık insan davranışlarının anlaşılmasını kolaylaştırmak ve etkileşimlerin kalitesini artırmak için kullanılmaktadır [65].

3.2. His Modelleri

Temel hislerin sınıflandırılması için çeşitli modeller geliştirilmiştir. Geliştirilen bu his modelleri kategorik, boyutsal ve genişletilmiş his modelleri olarak üç ana kategori altında incelenebilir [66].

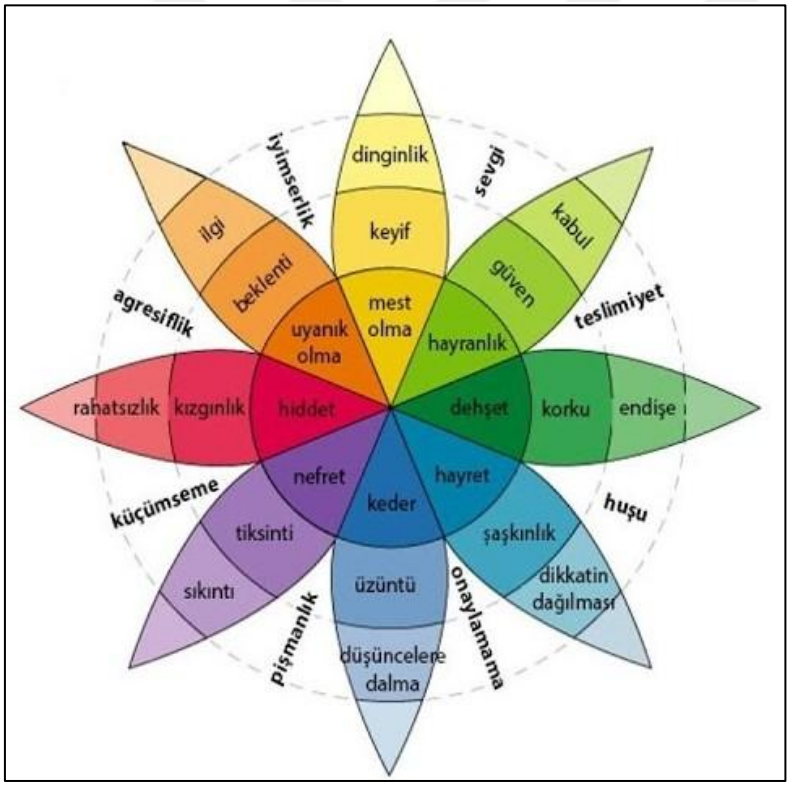
3.2.1. Kategorik his modelleri

Birey tarafından deneyimlenebilecek sınırlı bir his kümesi olduğu varsayımına dayanmaktadır [54]. Bu his modellerinde hisler ayrık biçimdedir. Yani bir bireyin aynı anda yalnızca bir hissi deneyimleyebileceği varsayılır. Paul Ekman'ın modeli his analizinde en sık kullanılan modellerden biridir. Ekman'ın modeline göre hisler sevinç, üzüntü, korku, öfke, tiksinti ve şaşkınlık olarak sınıflandırılır [54].

Şekil 3.2. Ekman his modeli [2]

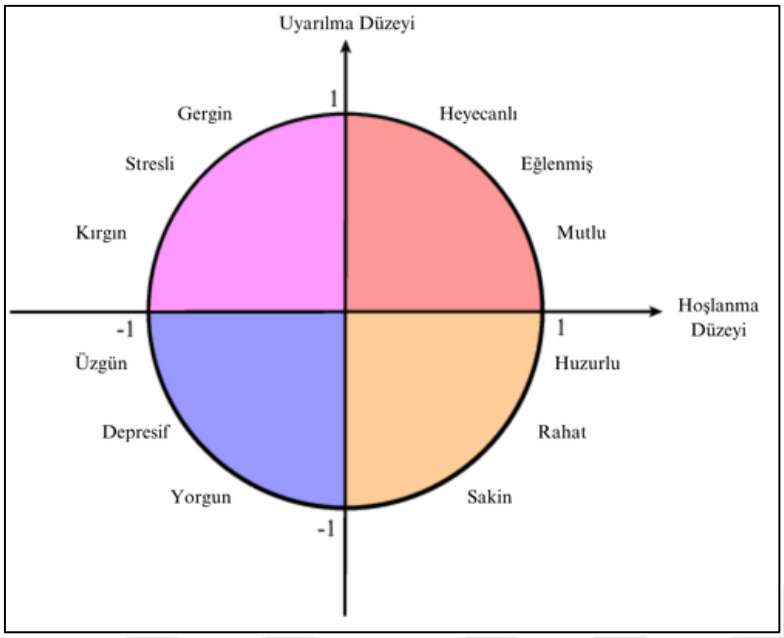
3.2.2. Boyutsal his modelleri

Bu his modeli hisleri tanımlamak için belirli faktörlere dayanan birden fazla boyut kullanır [67, 68]. Bu modeldeki boyutlar hissin negatif mi pozitif mi olduğunu ölçen valans, harekete geçirici mi yoksa sakin mi olduğunu ölçen uyarılma düzeyi ve hissin yoğunluğunu ölçen şiddettir [69]. Hisler bu temel kriterler göz önünde bulundurularak sınıflandırılır. Boyutsal his modelleri arasında yer alan Plutchik'in iki boyutlu his çarkı Şekil 3.3'te gösterilmiştir. Bu modelde yatayda bir uyarılma düzeyi ekseni dikeyde ise bir valans ekseni yer almaktadır. Metin tabanlı his analizi çalışmalarında en sık karşılaşılan sekiz temel duygu neşe, güven, korku, şaşırma, üzüntü, tiksinti, öfke ve beklenti olarak karşımıza çıkmaktadır.



Şekil 3.3. Plutchik his çarkı [2]

His analizi çalışmalarında kullanılan bir diğer model ise Russell'in his modelidir [66]. Bu model ise Şekil 3.4'te gösterildiği gibi uyarılma, valans ve şiddet boyutlarını kullanan üç boyutlu bir yaklaşım benimsemiştir.



Şekil 3.4. Russell his modeli [52]

3.2.3. Genişletilmiş his modelleri

Boyutsal ve kategorik his modellerine ek olarak metin tabanlı his analizi çalışmalarında Ortony, Clore ve Collins modeli ile Beş Faktör Kişilik Modeli tercih edilmektedir [70, 71]. Bu modeller daha farklı bir bakış açısı sunarak bireyin kişilik özellikleri, hedef ve arzuları gibi kişisel faktörleri dikkate alan çalışmalarda tercih edilerek his analizi sürecinin kişiye özgü hale getirilmesi amaçlamaktadırlar [66].

3.3. His Analizinde Kullanılan Veri Setleri

His analizinde kullanılan veri setleri doğru tespit ve sınıflandırma açısından önem taşır. Haber kaynakları, diyaloglar, dizi ve filmler, sosyal medya platformları gibi kaynaklardan elde edilen ve farklı dillerde sunulan veri setleri his analizinde kritik rol oynamaktadır. Çizelge 3.1'de his analizi çalışmalarında en sık karşılaşılan veri setlerinin temel özellikleri özetlenmiştir. Bu tez çalışmasında kullanılan DailyDialog, MELD ve IEMOCAP veri setleri üzerinde odaklanılmış olup, bu veri setleri ile ilgili özellikler metot ve materyal bölümünde açıklanmıştır.

Çizelge 3.1. His analizinde kullanılan veri setleri

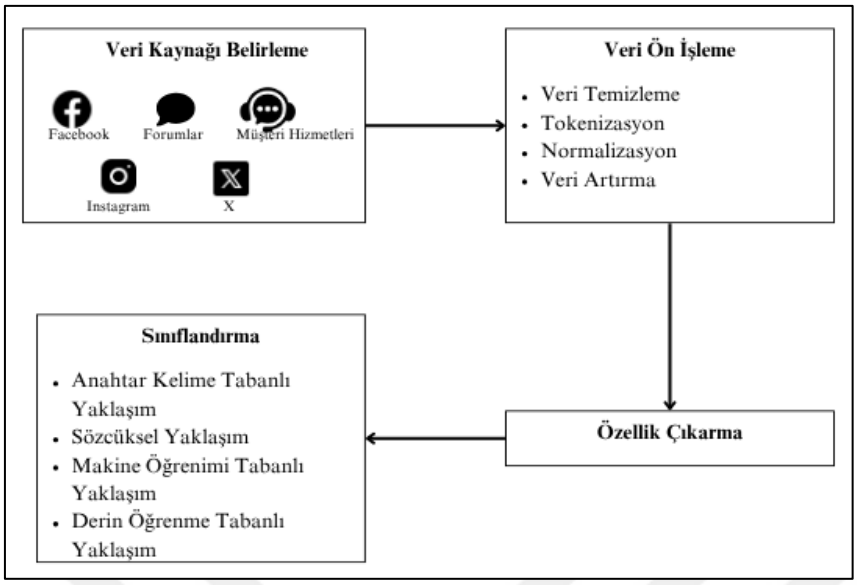
Veri Seti	Dil	Veri Türü	Etiketleme Türü	Veri Kaynağı	Çok Modelli Mi	Veri Büyüklüğü	His Modeli
SemEval (EmoInt-2007) [72]	İngilizce	Haber Başlıkları	Kategorik	Haberler	Hayır	1250	Ekman His Modeli
SemEval (EmoInt-2017) [73]	İngilizce	Kısa Metinler	Kategorik	Yarışma	Hayır	7000	Plutchik His Modeli
SemEval 2019 (EmoContext) [41]	İngilizce	Diyaloglar	Kategorik	Diyaloglar	Hayır	30 000	Ekman His Modeli
EMit (Emotions in Italian) [74]	İtalyanca	Kısa Metinler	Kategorik	Metinler	Hayır	1000	Ekman His Modeli
ISEAR [75]	Çok Dilli	Duygu Nedenleri	Duygu-Neden Eşleşmesi	Anket	Hayır	7625	Ekman-Izard His Modeli
EMOBANK [76]	İngilizce	Terimler Sözlüğü	Boyutsal	Sözlük	Hayır	10 000	VAD Model
EmoInt [73]	İngilizce	Kısa Metinler	Kategorik	Yarışma	Hayır	7000	Ekman His Modeli
ALM [77]	İngilizce	Hikayeler	Kategorik	Hikayeler	Hayır	188	Ekman His Modeli
IEMOCAP [78]	İngilizce	Diyaloglar, Hareketler	Kategorik	Diyaloglar	Evet	10 000	Ekman His Modeli
Sina Social News [79]	Çince	Haberler	Kategorik	Sosyal Medya	Hayır	3500	Ekman His Modeli
GoEmotions [80]	İngilizce	Haberler	Kategorik	Sosyal Medya	Hayır	58 000	Plutchik His Modeli
Ren-CECps corpus [81]	İngilizce	Bloglar	Kategorik	Blog	Hayır	1500	Ekman His Modeli
DailyDialog [82]	İngilizce	Günlük Diyaloglar	Kategorik	Diyaloglar	Hayır	13 000	Ekman His Modeli
CrowdFlower [83]	İngilizce	Soryal Medya, Forumlar	Kategorik	Sosyal Medya	Hayır	40 000	Ekman His Modeli
TEC [84]	İngilizce	Tweetler	Kategorik	Twitter	Hayır	21 000	Plutchik His Modeli
Grounded-Emotions [85]	İngilizce	Kısa Metinler	Kategorik	Twitter	Hayır	2557	İki Sınıflı
Emotion-Cause [86]	İngilizce	Duygu Nedenleri	Duygu-Neden Eşleşmesi	Çeşitli	Hayır	2500	Duygu-Neden Teorisi
SSEC [87]	İngilizce	Cümleler	Kategorik	Çeşitli	Hayır	11 855	Ekman His Modeli
TREMO [88]	İngilizce	Metinler	Kategorik	Çeşitli	Hayır	3000	Plutchik His Modeli

Çizelge 3.1. (devam) His analizinde kullanılan veri setleri

EmotionLines (Friends & EmotionPush) [89]	İngilizce	TV Diyalogları, Mesajlar	Kategorik	TV dizileri, Mesajlar	Hayır	25 000	Ekman His Modeli
TURTED [90]	Türkçe	Tweetler	Kategorik	Twitter	Hayır	20 000	Ekman His Modeli
EACWT [91]	Çince	Sosyal Medya Metinleri	Kategorik	Weibo	Hayır	5000	Plutchik His Modeli
Twitter, ROCStories [92]	İngilizce	Tweetler, Hikayeler	Kategorik	Sosyal Medya	Hayır	3000	Ekman His Modeli
Indonesian-language Tweets [93]	Endonezyaca	Tweetler	Kategorik	Twitter	Hayır	4403	Ekman His Modeli
MELD [94]	İngilizce	Diyaloglar, Video, Ses	Kategorik	TV Dizileri	Evet	1400 Diyalog, 13 000 Cümle	Ekman His Modeli
Readers' Emotion News (REN-20K) [21]	İngilizce	Haber Makaleleri	Kategorik	Web Sitesi	Hayır	20 474	Plutchik His Modeli
RENh-4K [95]	İngilizce	Haber Makaleleri	Kategorik	Web Sitesi	Hayır	4 000	Ekman His Modeli
Tales-Emotions [96]	İngilizce	Kısa Metinler	Kategorik	Masallar	Hayır	15 000	Ekman His Modeli
Electoral-Tweets [97]	İngilizce	Kısa Metinler	Kategorik	Twitter	Hayır	170 000	Plutchik His Modeli
EmoryNLP [98]	İngilizce	TV Diyalogları	Kategorik	TV dizisi	Hayır	12 606	Willcox'un Duygu Modeli

3.4. His Analizi Süreçleri

Metin tabanlı his analizi için bir takım adımların gerçekleştirilmesi gerekmektedir. Şekil 3.5'te his analizi sürecinin genel aşamaları gösterilmiştir.



Şekil 3.5. His analizinin genel aşamaları

3.4.1. Veri ön işleme

Diğer DDİ görevlerinde de olduğu gibi metin tabanlı his analizinde de veri ön işleme adımları modelin genel performansı açısından oldukça önemlidir. Veri ön işleme, modelin başarısını doğrudan etkileyen temel aşamalardan biridir [99]. Ham veriler, istenmeyen öznitelikler içerebileceğinden veya sınıf dengesizliği gibi problemlere yol açabileceğinden kapsamlı bir ön işleme sürecinden geçmesi gerekmektedir. Bu aşama bir dizi adımdan oluşur.

Veri temizleme

Veri temizleme adımı modelin başarısını olumsuz yönde etkileyecek, analiz açısından gerekli olmayan veya zararlı olabilecek verilerin kaldırılması işlemidir. Veri kalitesinin artırılması için oldukça önemli bir adımdır. Öncelikle noktalama işaretlerinin metinden kaldırılması ve tüm karakterlerin tek bir harf biçimine dönüştürülerek metnin standart hale getirilmesiyle veri kümesinde tutarlılık sağlanarak modelin dilsel varyasyonlardan etkilenmesi engellenir. Daha sonra bağlaçlar, edatlar gibi metin madenciliği açısından düşük öneme sahip ve belirli bir konuya özgü bilgi taşımayan durdurma sözcükleri metinden çıkarılır [100]. Özellikle sosyal medyadan elde edilmiş veri kümelerinde sıklıkla karşılaşılabilen kullanıcı etiketleri gibi içeriğin anlamına katkıda bulunmayan ögeler de temizlenmelidir. Sosyal medya kullanıcıları tarafından sıklıkla başvurulan emojiler ise

duygusal anlamın yakalanması için önemlidirler ancak mevcut DDİ araçlarının sınırlı bir bölümünün emoji tespit edip karşılık gelen anlamlarına dönüştürme yeteneğine sahip olduğu göz önünde bulundurulmalıdır [101]. Yine özellikle sosyal medya verilerinde gözlemlenen tekrarlayan karakterler, metin işleme algoritmaları tarafından farklı kelimeler olarak algılanabileceğinden bu karakterlerin sadeleştirilmesi daha doğru sonuçlar üretilmesine katkı sağlar [102].

Tokenizasyon

Tokenizasyon metnin anlamlı birimlere ayrıştırılmasını ifade eder [100]. Anlam tespit sürecinin ilk aşaması olan tokenizasyonun amacı karakterleri birbirinden ayırmak ve her bileşeni ayrı bir anlam birimi olarak tanımlamaktır [103]. Bu adımda metin semboller, kelimeler gibi temel bileşenlerine ayrılır [104].

Normalizasyon

Bu adımda metin üzerinde normalizasyon işlemi uygulanır. Normalizasyonun kök bulma ve lematizasyon olmak üzere iki temel yaklaşımı bulunmaktadır. Kök bulma işlemi kelimelerin temel hallerine indirgenmesidir. Örneğin “koşmak”, “koşucu”, “koşu” kelimeleri “koş” kökünden türetilmiştir. Ancak kök bulma işlemi sonucunda ortaya çıkan kök biçimler anlamsız kelimelere dönüşebildiğinden belirli dil bilgisi kurallarına uygun şekilde sadeleştirme konusunda sınırlı kalabilir ve kelimenin öz anlamını koruyamayabilir. Lematizasyon ise kelimenin anlamını koruyarak daha hassas bir çözüm sunar. Bu işlem kelimenin bağlamı ve türü dikkate alınarak uygulanır ve öncesinde tanımlanan sözcükler ile karşılaştırılarak uygun kök belirlenir. Geleneksel doğal dil işleme sistemlerinde bu adımlar ön işleme aşamasının bir parçasıdır. Ancak LLaMa gibi GDM’ler bağlamı anlamlandırabilme ve anlamsal ayrıştırma yeteneğine sahip olduğundan bu dönüşümlere ihtiyaç duyulmamaktadır.

Veri artırma

Derin öğrenme yöntemleri uygulanırken eğitim ve test aşamalarında büyük miktarda veriye ihtiyaç duyulmaktadır [105, 106]. Verinin yetersiz olduğu durumlarda ise veri artırma tekniklerine başvurulmaktadır [107]. Veri artırmanın temel amacı, modelin

genelleme kapasitesini artırmak için çeşitli tekniklerle veriyi yapay olarak artırmaktır. Özellikle his analizi gibi karmaşık görevlerde, artırılmış veri çeşitliliği modelin öğrenme sürecini güçlendirerek daha gerçekçi ve etkili çıktılar üretilmesine olanak tanımaktadır.

Geri çeviri, kelime gömme, eş anlamlı kelime değiştirme, metin üretimi gibi yöntemler yaygın veri artırma teknikleri arasında yer alır. Örneğin geri çeviri yöntemi bir metnin önce başka bir dile çevrilip sonrasında orijinal dile çevrilmesiyle farklı cümle yapılarının elde edilmesini sağlar. Bu işlem metnin anlamını korurken yapısal olarak çeşitlilik kazandırarak daha zengin bir veri kümesi oluşturur. Yapılan çalışmalar geri çevirinin dil modeli çeşitliliğini artırarak his analizi ve metin sınıflandırması gibi uygulamalarda performansı iyileştirdiğini göstermektedir [108, 109]. Kelime gömme ve eş anlamlı kelime değiştirme yöntemleri ise veri setini genişletirken, sahte haber tespiti ve his analizi gibi görevlerde önemli katkılar sağlamaktadır [110, 111]. Ayrıca, metin üretimi tekniklerinin hem kısa hem de uzun metin sınıflandırmalarında GDM'lerin gücünden faydalanarak etkili sonuçlar verdiği bilinmektedir [107, 112]. Veri artırma teknikleri arasında dikkat çeken bir diğer yöntem ise üretici rekabetçi ağlar (ÜRA)'dır. ÜRA mimarisi, iki yapay sinir ağı arasında rekabetçi bir yaklaşım benimseyerek, gerçek verilere benzer özellikler taşıyan yeni örnekler üreterek hem veri setini zenginleştirir hem de modelin daha iyi genelleme yapmasını sağlar [113].

3.4.2. Özellik çıkarımı

Özellik çıkarımı ham verilerin model açısından daha anlamlı hale getirilebilmesi için bir dizi özelliğe dönüştürülmesi sürecini ifade etmektedir. Veriler matematiksel bir model veya algoritma aracılığıyla yapılandırılmış bir forma dönüştürülür. Bu algoritmaların kullanımı metinsel verilerin vektör veya matris gibi sayısal temsillere dönüştürülmesini sağlayarak model tarafından analiz edilebilir bir forma çevrilebilmesini sağlar. Özellik çıkarma yöntemleri geleneksel ve modern yaklaşım olmak üzere iki ana kategoriye ayrılır.

<u>Geleneksel yaklaşımlar</u>

One-hot encoding, metinsel ifadeleri sayısal bir forma dönüştürmek için kullanılan temel bir tekniktir. Bu yöntemde sözlükteki her kelime konumuna karşılık gelen ikili vektörle temsil edilir. Vektörün yalnızca bir elemanı 1 diğer elemanları 0 olur. Örneğin elma

kelimesi [1,0] armut kelimesi [0,1] olarak gösterilir. Poslavskaya ve arkadaşları [114] çalışmalarında bu yöntemin özellikle çok sınıflı sınıflandırma problemlerinde etkili olduğunu belirtmişlerdir.

Kelimeler torbası yönteminde ise, bir metindeki kelimelerin frekansları sayılarak bir vektör oluşturulur. Her kelime, sözlükteki karşılığına göre 1 (var) veya 0 (yok) değeriyle temsil edilir. Kelimeler torbası yaklaşımı, kelimelerin cümledeki sırasını göz ardı eder; yalnızca kelimelerin tekrar sıklığına odaklanır. Aslam ve arkadaşları [115], bu yöntemi his analizi görevlerinde değerlendirerek, derin öğrenme tabanlı modellerle karşılaştırmalı analizler yapmışlardır.

Terim frekansı-ters belge frekansı, bir kelimenin belirli bir belge içerisindeki önemini, o kelimenin tüm belgeler üzerindeki yaygınlığına göre değerlendirir. Bu teknikte, terim frekansı terimin belgede kaç kez geçtiğini ölçerken, ters belge frekansı ise bu terimin tüm belgelerdeki göreli önemini belirlemektedir. Sık kullanılan kelimeler daha düşük ağırlık alır, nadir görülen kelimeler ise daha yüksek değerlerle temsil edilir. Qaiser ve arkadaşları [116] çalışmalarında bu yönteme odaklanmışlardır.

Sözcük türü etiketleme yönteminde ise metin içerisindeki her kelimenin dil bilgisel işlevi belirlenerek kelimeler isim, fiil, sıfat, zarf gibi türlere ayrılarak etiketlenir. Bu yöntem metin içerisindeki yapısal ilişkilerin analiz edilmesini kolaylaştırarak anlamsal çözümleme, söz dizimi çözümlemesi ve kelime anlamı belirleme görevlerinde önemli rol oynar.

Varlık adı tanıma yönteminde metin içerisinde yer alan özel isimler ve anlamlı varlıklar tespit edilerek kişi adları, kurumlar, tarihler, yerler, saatler, organizasyonlar gibi kategoriler altında sınıflandırılır. Bu yöntem bilgi çıkarımı, belge sınıflandırma, metin madenciliği gibi alanlarda kullanılmaktadır.

Modern yaklaşımlar

Geleneksel yöntemin sınırlılıkları nedeniyle araştırmacılar gömülü kelime temsili yöntemlerini önermişlerdir [117]. Gömülü kelime temsili yöntemleri geleneksel yöntemlerden farklı olarak bağlama uygun anlam yakalama, öğrenme kapasitesini artırma ve daha zengin temsiller sağlama avantajlarından dolayı tercih edilmektedir. Klasik

yöntemlerde kelimeler sabit vektörlerle temsil edilirken derin öğrenme tabanlı yaklaşımlarda kelime anlamı ve bağlamı da dikkate alınarak dinamik vektör temsilleri üretilir. Üretilen dinamik vektör temsilleri kelimeler arasında anlamlı ilişkiler kurulmasına yardımcı olarak büyük veri kümeleriyle eğitildiklerinde yüksek başarı sağlamaları bu yöntemlerin yaygınlaşmasını sağlamıştır.

Word2Vec yöntemi kelimelerin anlamlarını koruyacak şekilde N boyutlu vektörler halinde temsil eder. Mikalov ve arkadaşları [58] tarafından geliştirilen bu yöntem sürekli kelime torbası ve atlamalı n-gram olmak üzere iki temel yaklaşım içermektedir. Sürekli kelime torbası yaklaşımında belirli bir kelimenin çevresindeki kelimeler dikkate alınarak kelime tahmini yapılır. Atlamalı n-gram yaklaşımında ise belirli bir kelime üzerinden çevresindeki kelimeler tahmin edilir. Nandwani ve arkadaşları [118], sosyal medya ve çeşitli metinsel kaynaklardan türetilmiş duygu ve his analizleri üzerine yaptıkları çalışmada, Word2Vec yöntemini derin öğrenme teknikleriyle birleştirerek duyguların sınıflandırılmasını iyileştirmişlerdir. Buna ek olarak, Uymaz ve arkadaşları [119], Word2Vec, GloVe ve diğer yaygın kelime vektörleme yöntemlerini duygu analizi bağlamında karşılaştırmalı olarak incelemiş ve çalışmalarında Word2Vec'in bağlamsal anlamı koruma konusundaki avantajlarını vurgulamışlardır.

Glove ise kelimeleri N boyutlu vektörlerle ifade eden bir diğer gömülü temsildir [120]. Bu model büyük metin koleksiyonlarından elde edilen bir kelime matrisini kullanır. Kelimelerin kullandığı bağlamı analiz ederek kelimeler arasındaki ilişkiler vektörler ile temsil edilir. Benzer anlamlara sahip kelimeler benzer vektör temsilleriyle temsil edilmiş olur ve böylece kelimeler arasındaki benzerlikler ve anlamsal ilişkilerin yakalanmasını sağlar.

FastText kelimeleri karakter düzeyinde alt birimlere ayırarak temsil eden Facebook tarafından geliştirilmiş bir modeldir [121]. Benzer yazıma sahip kelimeler benzer vektörlerle ifade edilir. Bu yaklaşım özellikle nadir ve yeni kelimelerle başa çıkmada avantaj sağlayarak gömülü temsili daha etkili hale getirmektedir. Batbaatar ve arkadaşları [22], Word2Vec, GloVe ve FastText kullanarak duygusal özellikleri çıkarmış ve kelimelerle duygular arasındaki anlamsal ilişkileri başarıyla yansıtmışlardır.

ELMo tekniğinde ise bir terimin anlamsal değeri yalnızca kendisiyle değil çevresindeki

kelimelerle birlikte değerlendirilir. Çift yönlü UKSB ağı kullanılarak kelimenin içinde bulunduğu bağlama göre daha zengin bir temsil üretilir. Sarbazi-Azad ve arkadaşları [122], Arapça tweetlerde çoklu duygu tespiti amacıyla ELMo temsilleri ile UKSB modelini birleştirmiş ve bağlamsal anlamı koruma açısından yüksek başarı elde ettiklerini göstermişlerdir.

3.4.3. His analizinde kullanılan metotlar

Metin tabanlı his analizi için en yaygın kullanılan yöntemler anahtar kelimeye dayalı, sözcük temelli, makine öğrenimi temelli ve derin öğrenme temelli yaklaşımları kapsamaktadır.

Anahtar kelime tabanlı yaklaşımlar

Bu yaklaşım metin içerisindeki belirli sözcük veya sözcük gruplarını kullanmakta ve bu kelimelerin kullanım sıklığını ve temel kullanım bağlamını inceleyerek metindeki hissin tespit edilmesini amaçlamaktadır. Ancak bu yaklaşım sözcük bağlamını yüzeysel olarak dikkate aldığından daha karmaşık ve dolaylı ifadeleri doğru şekilde tespit edememe riski taşımaktadır.

Sözcük tabanlı yaklaşımlar

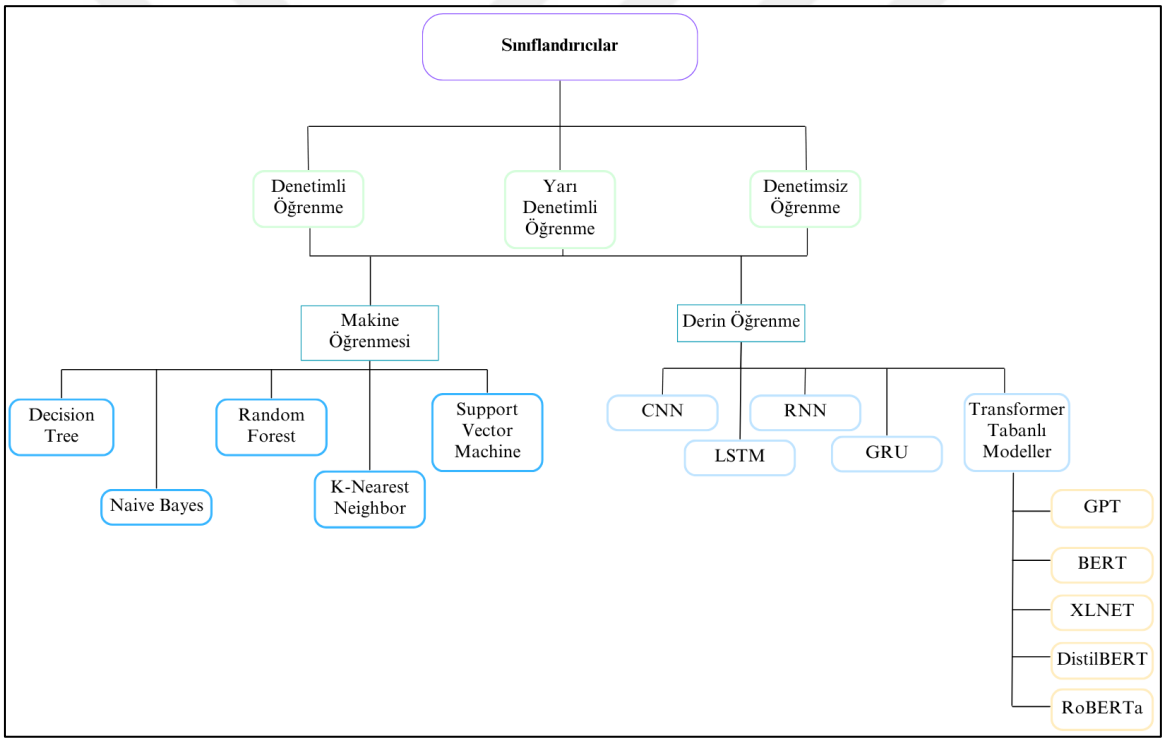
Genellikle metin analizi ve his tespiti gibi uygulamalarda kullanılan bu yaklaşım sözcük ve deyimlerin belirli bağlamlarda anlam ve çağrışımlarının incelenmesine dayanmaktadır. Metin içerisindeki belirli sözcüklerin frekansı ve taşıdığı anlam üzerinden genel duygusal durumu belirlenmektedir. Fakat dolaylı ifadeler barındıran metinlerde hissi tespit etmede sınırlılıkları bulunmaktadır.

Makine öğrenmesi tabanlı yaklaşımlar

Makine öğrenmesi tabanlı yaklaşımlarda metindeki hissi doğrudan kelime eşleştirmesi ile belirlemek yerine daha önce öğrenilen örüntülere dayanarak yeni metinlerdeki hissi tahmin eden bir model oluşturulur. Bu yaklaşım denetimli öğrenme ve denetimsiz öğrenme olmak üzere iki temel kategoriye ayrılmaktadır. Denetimli öğrenmede etiketlenmiş veri kümeleri

kullanılarak metin özellikleri öğrenilir ve tahmin yapılır. Denetimsiz öğrenmede ise denetimli öğrenmeden farklı olarak etiketlenmemiş veriler analiz edilerek verilerdeki gizli kalıp ve örüntülerin ortaya çıkarılması amaçlanır. Model, verilerin içinde hangi hislerin olabileceğini doğrudan öğrenmez. Bunun yerine, benzer içeriklere sahip metinleri gruplandırarak aralarındaki ilişkiyi öğrenmeye çalışır.

Bu çalışmada geleneksel makine öğrenimi yöntemleri yerine derin öğrenmeye dayalı yöntemlerden dönüştürücü tabanlı GDM'lere odaklanıldığından geleneksel makine öğrenimi yöntemlerinin ayrıntılarına yer verilmemiştir. Metin tabanlı his analizinde kullanılan sınıflandırıcılar Şekil 3.6'da yer almaktadır.



Şekil 3.6. Metin tabanlı his analizinde kullanılan sınıflandırıcılara genel bakış [123]

Derin öğrenme tabanlı yaklaşımlar

Derin öğrenme, makine öğreniminin bir alt dalını oluşturmaktadır. Kullandığı hiyerarşik yapı ile işlenmemiş verilerden karmaşık özelliklerin otomatik olarak keşfedilmesini kolaylaştırır. Çok katmanlı bir yaklaşım benimseyen derin öğrenme yöntemlerinin uygulanması belirli alanlar için yeni sınıflandırıcı ve araçlar geliştirilmesini kolaylaştırarak metin sınıflandırma görevlerinin etkinliğini artırmaktadır. Denetimli öğrenme, denetimsiz

öğrenme, ve yarı denetimli öğrenme senaryolarında etkili şekilde kullanılabilen derin öğrenme sınıflandırıcıları öğrenme yetenekleri sayesinde hem doğruluk hem de genel performans açısından önemli iyileştirmeler sağlamaktadır. Literatürde yaygın olarak kullanılan derin öğrenme sınıflandırıcıları arasında ESA, YSA, UKSB, GYB, BERT ve çeşitli önceden eğitilmiş modeller yer almaktadır. Bu yapılar, farklı doğal dil işleme görevlerinde yüksek doğruluk oranları ile başarılı sonuçlar elde edilmesini mümkün kılmaktadır. Bu yöntemler arasında özellikle son yıllarda dönüştürücü tabanlı mimari üzerine inşa edilen modeller dikkat çekmektedir. Dikkat mekanizması ve çok katmanlı yapısıyla bağlamsal bilgi işleme konusunda yüksek başarı sağlayan bu modeller, özellikle GDM'lerin geliştirilmesinde temel oluşturmaktadır. BERT, RoBERTa, GPT ve LLaMA gibi GDM'ler, transformer tabanlı mimariyi kullanmaktadır. Dönüştürücü tabanlı modellerden LLaMA'nın mimari detayları ve bu çalışmadaki uygulama biçimi ilerleyen bölümlerde ayrıntılı şekilde ele alınmaktadır.

4. METOT VE MATERYAL

Bu bölümde, çalışmada kullanılan veri kümeleri, ön işleme adımları, model yapılandırmaları ve deneysel ayarlar ayrıntılı olarak sunulmuştur. GDM'lerle gerçekleştirilen sıfır atışlı ve az atışlı öğrenme yaklaşımları ile ince ayar süreçleri açıklanmıştır. Kullanılan yöntemlerin karşılaştırmalı analizi, her bir model yapılandırmasının sistematik olarak değerlendirilmesine olanak tanımıştır.

4.1. Veri Seti ve Veri Ön İşleme

Çalışmada veri kümesi olarak DailyDialog, MELD ve IEMOCAP kullanılmıştır. Kullanılan veri setileri ile ilgili istatistikler Çizelge 4.1 ve Çizelge 4.2'de yer almaktadır. Veri setleri ile ilgili ayrıntılar ise aşağıda açıklanmıştır.

Çizelge 4.1. Veri setlerinin diyalog ve konuşma satırı sayıları (eğitim / doğrulama / test)

	DailyDialog	MELD	IEMOCAP
#Diyalog	2403	1432	151
Eğitim	403	1038	100
Doğrulama	1000	114	20
Test	1000	280	31
#Konuşma Satırı	30280	13708	7258
Eğitim	14471	9989	4690
Doğrulama	8069	1109	970
Test	7740	2610	1598
Sınıf Sayısı	7	7	6

Çizelge 4.2. Veri setlerinde his sınıfı dağılımı (eğitim / doğrulama / test)

	DailyDialog			MELD			IEMOCAP		
Duygu	Eğitim	Doğrulama	Test	Eğitim	Doğrulama	Test	Eğitim	Doğrulama	Test
nötr	7595	7108	6321	4710	470	1256	1167	156	382
mutluluk	3031	684	1019	1743	163	402	357	57	134
şaşkınlık	1600	107	118	1205	150	281	-	-	-
üzüntü	969	79	116	683	111	208	739	100	245
öfke	827	77	102	1109	153	345	711	222	170
iğrenme	303	3	47	271	22	68	-	-	-
korku	146	11	17	268	40	50	-	-	-
hayal kırıklığı	-	-	-	-	-	--	1146	319	380
heyecanlı	-	-	-	-	-		570	116	287

DailyDialog veriseti [82] günlük İngilizce diyalogları içerir ve günlük diyaloglardaki hislerin tespiti için kullanılır. Veri seti diyalogları ve bu diyaloglarla ilişkili his etiketlerini içerir. Ekman'ın tanımladığı altı temel duygu kategorisi olan öfke (anger), iğrenme (disgust), korku (fear), mutluluk (happiness), üzüntü (sadness) ve şaşkınlık (surprise) etiketlerinden biriyle etiketlenmiştir. Ayrıca herhangi bir duygu içermeyen ifadeler nötr (no emotion) olarak etiketlenmiştir.

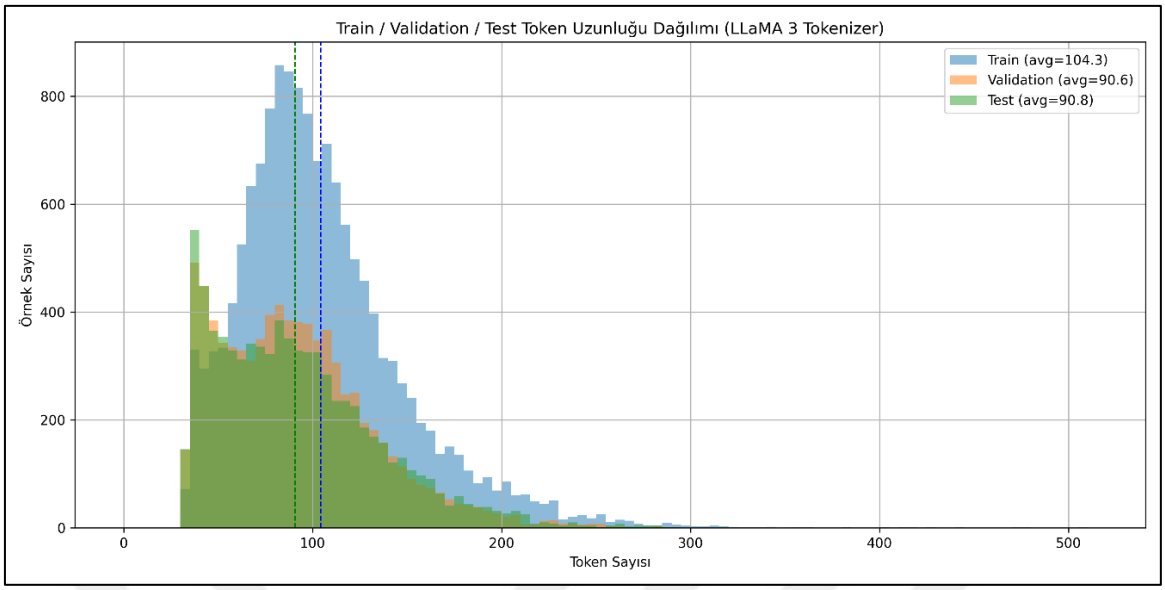
MELD veri seti [94], Friends televizyon dizisinden 1.400'den fazla konuşma ve 13.000'den fazla cümleye sahip ve içerisinde metin, ses ve video verilerini bulunduran İngilizce bir veri setidir. Bu çalışma kapsamında yalnızca metin verileri kullanılmıştır. Veri seti öfke (anger), iğrenme (disgust), korku (fear), mutluluk (happiness), üzüntü (sadness)**,** şaşkınlık (surprise) ve nötr (no emotion) olarak yedi farklı etiket ile etiketlenmiştir.

IEMOCAP veri seti [78], oyuncular tarafından gerçekleştirilen ikili konuşmalardan oluşmakta ve hem senaryo tabanlı hem de doğaçlama diyalogları içermektedir. Her bir konuşma ses, video ve metin verileriyle birlikte sunulmaktadır. Bu çalışmada yalnızca metin tabanlı içerikler kullanılmıştır. Bu çalışmada kullanılan his sınıfları, temel olarak Ekman'ın duygu modeline dayanan ancak konuşma temelli veri setlerinde daha sık

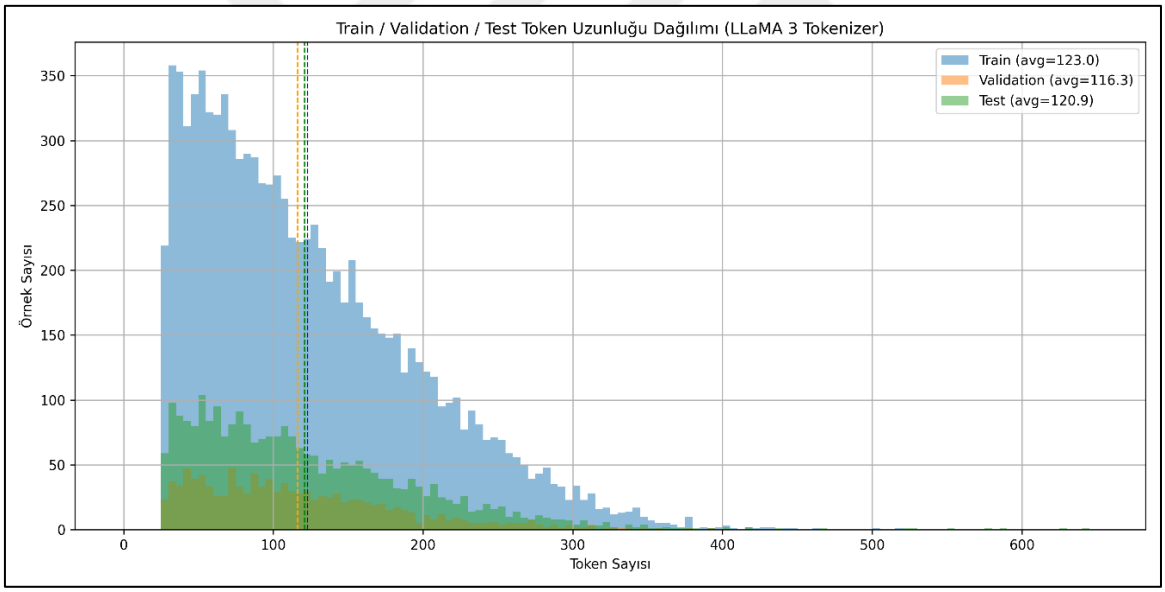
karşılaşılan durumları da kapsayacak şekilde genişletilmiş bir yapıdadır. Veri seti nötr (neutral), hayal kırıklığı (frustrated), öfke (angry), üzüntü (sad), heyecanlı (excited) ve mutluluk (happy) etiketleri ile etiketlenmiştir.

Uygulanan veri ön işleme süreci, çalışmada kullanılan Meta-LLaMA3 8B Instruct modelinin yapısına uygun olarak tasarlanmış ve geleneksel kodlayıcı tabanlı modellerden farklı olarak istem tabanlı bir yaklaşım benimsenmiştir. İlk adımda, veri setlerindeki Unicode karakter bozulmaları giderilmiş ve tüm metinler ASCII karakterlerine dönüştürülerek standartlaştırılmıştır. Her bir diyalog örneği, içerdiği ifade sayısına göre özgün formunda işlenmiştir. GDM sabit uzunlukta girişlere ihtiyaç duymadığından ve hem doğal diyalog akışını korumak hem de modele sahte bağlamlar verilmesini önlemek amacıyla ve herhangi bir padding işlemi uygulanmamıştır.

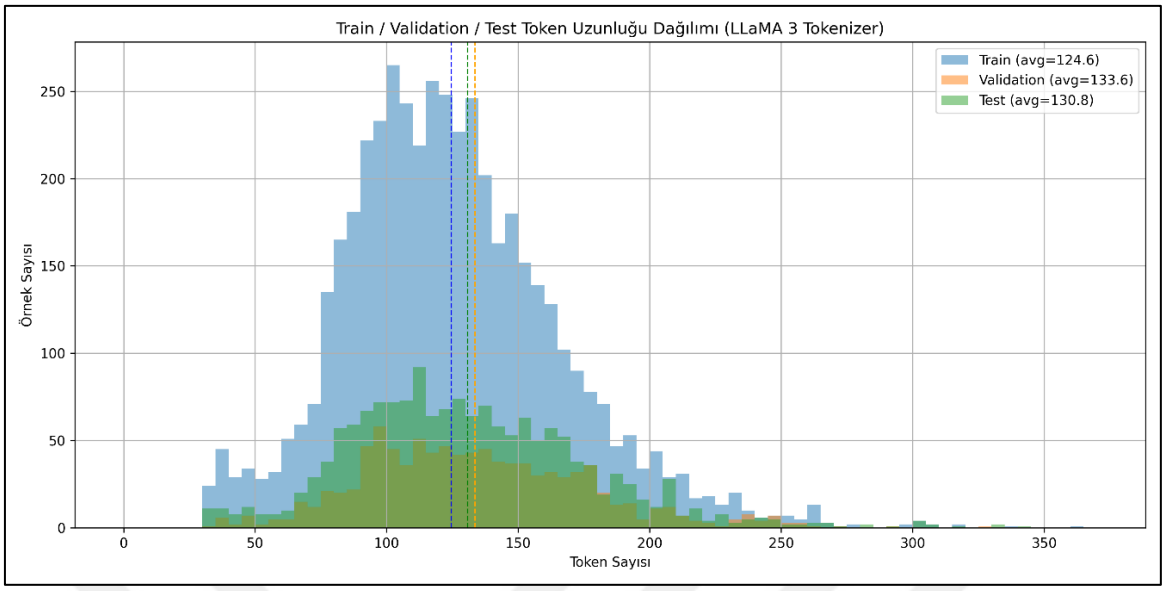
Modelin belleğini verimli kullanabilmek için, giriş ve çıkış metinlerinin uzunluğu çoğu örnekte olduğu gibi korunmuştur. Ancak, LLaMA 3 8B Instruct modelinin maksimum çıktı uzunluğu 2048 token ile sınırlı olduğundan**,** toplam uzunluğun bu sınırı aşabileceği durumlarda giriş metinleri dikkatli şekilde kısaltılmıştır. Bu işlem yalnızca gerekli olduğunda uygulanmış, aksi halde örnekler özgün halleriyle modele sunulmuştur. Şekil 4.1, 4.2 ve 4.3'te, veri setlerindeki örneklerin LLaMA 3 tokenizer tarafından ayrıştırılması sonucunda elde edilen ortalama token uzunlukları sunulmaktadır.



Şekil 4.1. DailyDialog veri seti token uzunluğu dağılımı (eğitim / doğrulama / test)



Şekil 4.2. MELD veri seti token uzunluğu dağılımı (eğitim / doğrulama / test)

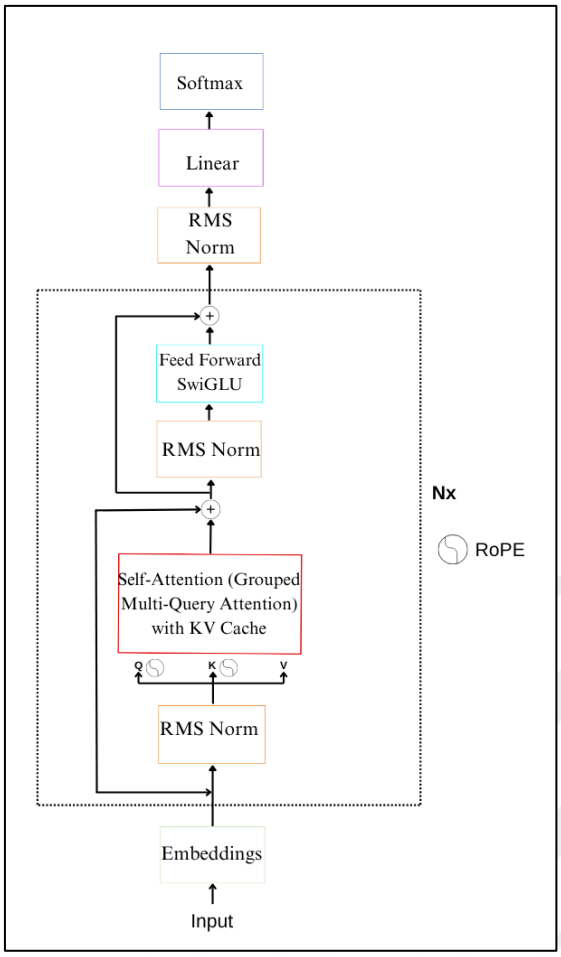


Şekil 4.3. IEMOCAP veri seti token uzunluğu dağılımı (eğitim / doğrulama / test)

Her örnek, modelin görevi doğru şekilde anlayabilmesi için yönlendirici bir sistem mesajı ile başlatılmıştır. Ardından, önceki konuşma satırları (bağlam) ve hedef ifade özel olarak yapılandırılmış bir metin formatında birleştirilerek modele sunulmuştur. Her bir örneği modele iletmek için oluşturulan istemin yapısı 4.3.1. İstem Yapısının Bileşenleri bölümünde açıklanmıştır.

4.2. Model Mimarisi ve Kullanım Yöntemleri

Meta AI tarafından geliştirilen ve dönüştürücü tabanlı mimari üzerine inşa edilmiş olan LLaMA 3 8B Instruct modeli, bu çalışmada kullanılan temel model olarak tercih edilmiştir. Modelin mimari tasarımı Şekil 4.4'te görselleştirilmiştir.



Şekil 4.4. LLaMa mimarisi [124]

Meta AI tarafından yayımlanan resmi rapora göre [124], model toplamda 32 katmandan oluşmaktadır ve her bir katmanda 4096 boyutlu bir gizli temsil alanı bulunmaktadır. Modeldeki katmanlar arasında bilgi aktarımını sağlayan çoklu dikkat mekanizması ve her katmanda 32 adet dikkat başlığı yer almaktadır. Ancak hesaplama yükünü azaltmak ve belleği daha verimli kullanmak amacıyla yalnızca 8 başlık anahtar-değer hesaplamaları için kullanılır. Bu tasarım gruplandırılmış sorgu dikkati adı verilen bir dikkat yöntemine dayanmaktadır. Bu yöntemde klasik çoklu dikkat mekanizmasındaki her başlık için ayrı anahtar-değer vektörleri üretmek yerine sadece bazı başlıklar için anahtar-değer hesaplanarak sorgu vektörlerinin bu hesaplamaları gruplar halinde paylaşmasına olanak tanınır. Modelin öğrenme sürecinde SwiGLU adlı özel bir aktivasyon fonksiyonu kullanılarak modelin karmaşık örüntüleri öğrenmesine yardımcı olunmuştur. Katmanlar arasında denge sağlamak amacıyla klasik katman normalizasyonu yerine RMS normalizasyonu kullanılmıştır. Modelin konum bilgisini anlayabilmesi için taban frekansı artırılarak modelin daha uzun metinlerle başa çıkabilme kapasitesini geliştirmek amacıyla

RoPE adlı yöntem kullanılmıştır. Modelde 128 000 farklı kelime veya kelime parçasını tanıyabilen özel bir tokenizer kullanılmıştır. LLaMa 3 modeline ait tüm bu mimari özellikler modeli bağlama duyarlı ve güçlü bir dil modeli haline getirmiştir. Ayrıca model açık kaynaklı olduğundan araştırma alanında yeniden üretilebilirliği desteklemektedir.

Mimari altyapının yanı sıra, bu çalışmada modelin sıfır atış öğrenme, az atış öğrenme, ince ayar kullanım biçimleri de ele alınarak çeşitli deneysel senaryolar geliştirilmiştir.

4.2.1. Sıfır atış öğrenme yaklaşımı

Sıfır atış öğrenme, modelin daha önce görmediği bir görev veya sınıf için yalnızca görev tanımı üzerinden yönlendirilerek tahmin yapabilmesini ifade eder. Bu yöntemde modele herhangi bir örnek gösterilmez. Model önceden eğitilmiş genel bilgileriyle ve verilen istem aracılığıyla sonuca ulaşır.

Bu çalışmada kullanılan sıfır atış yönteminin ilk aşaması, modeli doğru bir şekilde yönlendirmek için istemin oluşturulmasıdır. Kullanılan test örnekleri, ifadenin ortaya çıktığı bağlamı yansıtmak amacıyla, önceki konuşma satırlarını ve sınıflandırılacak son ifadeyi kapsamaktadır. Önceki konuşma satırlarının modele sunulması, anlamın doğru bir şekilde yorumlanması ve daha isabetli tahmin edilebilmesi için kritik öneme sahiptir. Modelin görevi daha doğru bir şekilde anlayabilmesi için her isteme sistem mesajı dahil edilmiştir. Modelin duyguları sınıflandırmadaki etkinliği, örneklerin sunulmasına gerek kalmaksızın, yalnızca görev açıklamalarına dayalı olarak bunu yapma yeteneği ile kanıtlanmıştır. Sıfır atış yaklaşımı için kullanılan istem yapısına istem tasarımı bölümünde yer verilmiştir.

4.2.2. Az atış öğrenme yaklaşımı

Az atışlı öğrenme yaklaşımında modelin görev hakkında daha iyi genelleme yapabilmesi için her sınıftan az sayıda örnek modele sunulur. Bu yaklaşımda sıfır atış yaklaşımından farklı olarak görev tanımıyla birlikte birkaç örnek gösterilerek modelin örüntüleri öğrenmesi ve benzer bağlamlarda doğru çıkarım yapabilmesi hedeflenir. Modelin parametreleri güncellenmez, yönlendirici istem üzerinden üretim yapması sağlanır. Az atış öğrenme yaklaşımında örnek seçimi modelin başarımı için kritik öneme sahiptir.

Bu çalışmada istem tasarımında sıfır atışta kullanılan isteme ek olarak her sınıftan bir örnek isteme dahil edilmiştir. Bu örnekler, modelin görev formatını öğrenmesi amacıyla sistem mesajından sonra sırayla yerleştirilmiş ve her biri bağlam, hedef ifade ve karşılık gelen duygu etiketini içermiştir. Test örneği ise bu örneklerden sonra sunularak modelden yalnızca son ifadede yer alan hissin tahmin edilmesi beklenmiştir. Kullanılan istem yapısına istem tasarımı bölümünde yer verilmiştir.

4.2.3. İnce ayar eğitimi

İnce ayar ise sıfır atışlı öğrenme ve az atışlı öğrenme yaklaşımlarının aksine önceden eğitilmiş modelin parametrelerinin belirli bir görev için sonradan etiketli verilerle yeniden eğitilmesi sürecini ifade eder. Bu yöntem modelin görevle ilgili bilgileri doğrudan öğrenmesini sağlayarak en yüksek başarıma ulaşmayı hedefler. Ancak bu süreçte ihtiyaç duyulan hesaplama gücü ve eğitim süresi sıfır atış öğrenme ve az atış öğrenme yaklaşımlarına göre çok daha yüksektir.

Yapılan çalışma kapsamında Meta LLaMA 3 8B Instruct modeli his analizine yönelik olarak etiketlenmiş verilerle yeniden eğitilmiştir. Yapılan eğitimde modelin tam olarak eğitilmesi yerine parametre etkin ince ayar (PEİA) yaklaşımı benimsenmiştir. Ayrıca düşük-ranklı uyumlama (DRU) yöntemi uygulanarak modele düşük boyutlu adaptasyon matrisleri entegre edilmiş ve yalnızca bu bölümler eğitilmiştir. Bunların yanı sıra model çıktısını his sınıflarına indirgeyebilmek için son katmana bir sınıflandırıcı eklenmiştir. Eğitim sürecinde model, görev tanımı içeren bir istem yapısıyla yönlendirilmiştir. Kullanılan bu istem yapısı ile ilgili ayrıntıya istem tasarımı bölümünde, kullanılan yaklaşım ve yöntemlere ise eğitim stratejisi ve optimizasyon bölümünde yer verilmiştir.

4.3. İstem Tasarımı

Çalışmada kullanılan LLaMA 3 8B Instruct modeli, doğal dil girdileri yardımıyla yönlendirilen bir GDM'dir. Bu nedenle modelin görev tanımını anlayabilmesi ve hedef çıktıyı doğru şekilde üretebilmesi için modele sunulan istemlerin dikkatli bir şekilde yapılandırılması gerekmektedir. Modele sunulan istemler başarımı etkilemektedir. Bu bölümde sıfır atış, az atış ve ince ayar senaryolarında kullanılan istem tasarımları ayrıntılı olarak açıklanmıştır.

4.3.1. İstem yapısının bileşenleri

Modelin yönlendirilmesi için kullanılan istemler genel olarak görev tanımı, diyalog geçmişi, hedef ifade ve çıktı yönlendirmesi olmak üzere dört ana bileşenden oluşmaktadır.

Sistem mesajı

Modelin bir his sınıflayıcı olarak davranmasını sağlamak amacıyla, senaryolarda istemler bir sistem mesajı ile başlatılmıştır. Bu mesajda modelin hangi rolü üstleneceği açıkça belirtilmiştir.

Görev tanımı

Tüm senaryolarda modelin görev çerçevesini anlayabilmesi amacıyla istemlerde görev tanımına yer verilmiştir. İnce ayar sürecinde bu tanım "### Instruction:" etiketiyle açıkça belirtilmişken sıfır atış ve az atış senaryolarında, sistem mesajı ve kullanıcı girdisi içinde doğal dil biçiminde modele sunulmuştur.

Diyalog geçmişi

Tüm senaryolarda, son ifadenin bağlamını oluşturacak şekilde önceki konuşma satırları "Previous conversation:" ya da "###Conversation History:" etiketiyle modele sunulmuştur.

Hedef ifade

Duygusu sınıflandırılmak istenen son ifade "Now the utterance is:" ya da "### Last Utterance:" etiketi ile belirtilmiştir.

Çıktı yönlendirmesi

Modelin yalnızca his etiketi üretmesini sağlamak amacıyla "Emotion:" ya da "### Response:" satırı eklenmiştir.

System: You are an expert emotion classifier. Respond with one of the following labels: no emotion, anger, disgust, fear, happiness, sadness, surprise

Previous conversation:

A: I can’t believe this is happening!

B: It’s been so stressful lately.

Now the utterance is:

I feel like I’m losing control.

Emotion:

Şekil 4.5. Sıfır atışlı öğrenme yaklaşımı istem tasarımı

System: You are an expert emotion classifier. Respond with one of the following labels: no emotion, anger, disgust, fear, happiness, sadness, surprise

Example 1:

Previous conversation:

A: I just got promoted today.

B: That’s amazing!

Now the utterance is:

I feel so proud and excited.

Emotion: joy

[Test instance]

Previous conversation:

A: I can’t believe this is happening!

B: It’s been so stressful lately.

Now the utterance is:

I feel like I’m losing control.

Emotion:

I feel so proud and excited.

Emotion: joy

[Test instance]

Previous conversation:

A: I can’t believe this is happening!

B: It’s been so stressful lately.

Now the utterance is:

I feel like I’m losing control.

Emotion:

Şekil 4.6. Az atışlı öğrenme yaklaşımı istem tasarımı

Instruction:Given a conversation history, determine the emotion expressed in the last utterance. Respond using ONLY one of the following labels: anger, disgust, fear, happiness, sadness, surprise, no emotion.

Conversation History:

A: Hey man , you wanna buy some weed ?

B: Some what ?

Last Utterance:

A: Weed ! You know ? Pot , Ganja , Mary Jane some chronic !

Response:

Şekil 4.7. İnce ayar öğrenme yaklaşımı istem tasarımı

4.3.2. Çıktı alanının sınırlandırılması

Sıfır atış ve az atış senaryolarında, istem yapısında "Emotion:" ifadesi kullanılarak modelden yalnızca belirli etiketlerden birini üretmesi istenmiş olsa da, bu yönlendirme tek başına modelin çıktısını kısıtlamak için yeterli olmamıştır. LLaMA 3 gibi GDM'ler, doğaları gereği serbest metin üretmeye eğilimlidir ve verilen istem doğrultusunda çok çeşitli, bazen beklentinin dışında ifadeler üretebilirler. Örneğin, modelin "Emotion: happy" gibi doğrudan etiket üretmesi gerekirken "I think it's happiness" gibi tamamlayıcı, açıklamalı metinler de üretebilmektedir. Bu durum, çıktıların hem karşılaştırılabilirliğini zorlaştırmakta hem de değerlendirme sürecinde ek normalizasyon adımlarını zorunlu kılmaktadır. Bu sorunu önlemek ve modelin yalnızca geçerli duygu etiketlerinden birini doğrudan üretmesini güvence altına almak amacıyla, çıkarım sürecine müdahale eden logit düzeyinde denetim mekanizması uygulanmıştır. Bu mekanizma, çıkarım aşamasında devreye giren ve modelin olası çıktılarını logit düzeyinde filtreleyen bir logit işleyici yapısıdır.

Logit işleyici, modelin her üretim adımında hesapladığı logit vektörünü $z = [z_1, z_2, ..., z_v] \in \mathbb{R}^v$ biçiminde ele alır. Burada *V*, modelin tüm kelime dağarcığını temsil etmektedir. Bu yapı, yalnızca önceden tanımlanmış geçerli etiketlere karşılık gelen token konumlarını koruyacak şekilde çalışır; diğer tüm token'ların logit değerlerini -∞ ile bastırır.

$$z_i = \begin{cases} z_i, & \text{eğer } i \in V_{geçerli} \\ -\infty, & \text{aksi takdirde} \end{cases} \quad (4.1.)$$

Bu bastırma işlemi, softmax fonksiyonu uygulanmadan önce geçersiz etiketlere ait olasılıkların sıfırlanmasını sağlar. Böylece modelin yalnızca sınıflandırma görevine uygun olan etiketlerden birini üretmesi zorunlu hale gelir. Bu strateji, modelin çıktı uzayını büyük ölçüde daraltarak, hem üretim kararlılığını artırmış hem de sıfır atışlı senaryolarda doğruluğu optimize etmiştir. Bu bağlamda, logit düzeyinde yapılan bu müdahale, açık uçlu üretim özelliklerine sahip dil modellerinin sınıflandırma görevlerine daha güvenilir biçimde adapte edilmesini sağlamış ve değerlendirme süreçlerinde normalizasyon ihtiyacını azaltmıştır. Uygulanan bu sınırlama sadece çıkarım sürecinde geçerlidir, ince ayar sürecinde model doğrudan etiketli veriyle eğitildiği ve çıktı üretimi sınıf etiketleriyle sınırlandığı için logits düzeyinde bir filtreleme ihtiyacı doğmamıştır.

4.4. Tokenizasyon ve Girdi Hazırlama Süreci

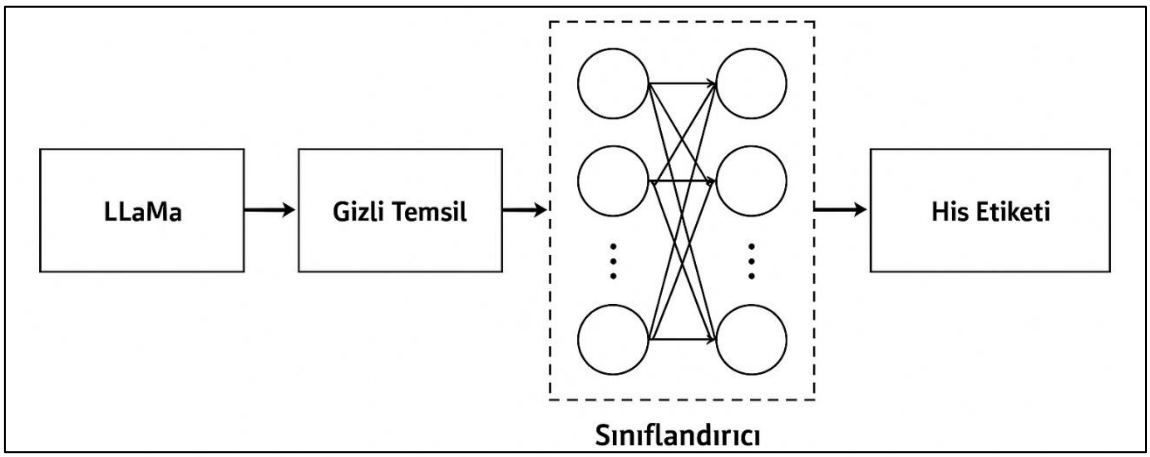
Dil modelleri, doğal dil girdilerini doğrudan işleyemediğinden metinlerin öncelikle daha küçük ve anlamlı birimlere ayrılarak sayısal temsillere dönüştürülmesi gerekmektedir. Modelin metni anlayabilmesi için bir ön koşul olan bu işlem tokenizasyon olarak adlandırılır.

Bu çalışmada, kullanılan Meta-LLaMA 3 8B Instruct modeline tam uyumlu olacak şekilde Meta tarafından geliştirilen ve önceden eğitilmiş bir AutoTokenizer kullanılmıştır. Bu tokenizer, modelin sözcük hazinesi ve kodlama şemasıyla birebir uyumlu şekilde çalışır ve alt birim temelli geniş bir sözcük dağarcığına sahiptir. Bu yapı, özellikle diyaloğa dayalı metinlerde karşılaşılan nadir, çok anlamlı veya biçimsel olarak karmaşık ifadelerin daha doğru biçimde modellenmesini sağlamaktadır.

Modelin teknik sınırlamaları göz önünde bulundurularak giriş dizileri belirli bir maksimum uzunlukla sınırlandırılmış ve gerektiğinde kırpma işlemi uygulanmıştır. Eğitim verileri, sabit uzunlukta olacak şekilde önceden tanımlı bir doldurma işlemiyle doldurulmamış; bunun yerine, batch işlemleme sırasında otomatik padding mekanizması devreye alınmıştır. Bu süreçte, LLaMA 3 modelinin kendi özel durdurucu (EOS) token'ı, tokenizer yapılandırmasında doldurucu token olarak tanımlanmıştır. Böylece, modelin dahili işlem mantığıyla tam uyum sağlanarak padding sırasında oluşabilecek anlamsal bozulmalar önlenmiş ve bağlamsal bütünlük korunmuştur. Bu yapılandırma aynı zamanda grafik işlem birimi (GİB) belleğinin verimli kullanımına da katkı sağlamaktadır.

4.5. İnce Ayar İçin Eğitim Stratejisi ve Optimizasyon

Yapılan çalışmada ince ayar yaklaşımı için, GDM'lerin his analizi görevine uyarlanmasında, sınırlı sayıda etiketli veriye rağmen yüksek genelleme başarımı elde edebilmek amacıyla çok katmanlı bir eğitim stratejisi izlenmiştir. Eğitim süreci, hesaplama verimliliği, sınıf dengesizliğiyle başa çıkma, görev odaklı çıkış katmanı inşası ve veri temsiliyetinin artırılması gibi temel hedefler doğrultusunda yapılandırılmıştır. İnce ayar için uygulanan eğitim stratejisi genel hatlarıyla Şekil 4.8'de gösterilmiştir.



Şekil 4.8. İnce ayar için eğitim stratejine genel bakış

4.5.1. Parametre verimli ince ayar yaklaşımı

Geleneksel tam model güncellemesine kıyasla çok daha az sayıda parametreyle yüksek başarı elde etmek için parametre verimli ince ayar yöntemlerinden biri olan DRU tercih edilmiştir. Bu yöntem, yalnızca modelin dikkat mekanizması içindeki belirli katmanlarına güncellemeler uygulayarak, modelin geri kalan kısmını sabit tutar. Böylece, eğitim sürecinde hem hesaplama kaynakları azaltılmış hem de önceden eğitilmiş dil bilgisi korunarak göreve özgü bilgi model tarafından daha etkin öğrenilmiştir. Ayrıca, modelin belleğe yükünü azaltmak amacıyla dörtlü bit kuantizasyon yöntemi kullanılmıştır. Bu yöntem, GDM'lerin daha küçük donanımlarda eğitilmesini mümkün kılarken, DRU ile birlikte uyumlu çalışarak hem parametre verimliliği hem de hesaplama verimliliği sağlamıştır.

4.5.2. Göreve özgü sınıflandırıcı katman tasarımı

LLaMA gibi GDM'ler, metin üretimi görevleri için optimize edildiğinden doğrudan his sınıflandırması yapabilecek bir çıkış yapısına sahip değildir. Bu nedenle, modelin ürettiği gizli temsilleri alıp sınıf etiketlerine dönüştürebilen görev odaklı sınıflandırıcı katman geliştirilmiştir. Bu sınıflandırıcı, modelin çıktısındaki son gizli katmandan, bağlamı en iyi yansıtan ve genellikle tüm diziyi temsil etme yeteneğine sahip olan özel bir token üzerinden özet bir temsil elde ederek bunu doğrusal katmanlar üzerinden geçirerek tahmin yapar. Bu yapı, görevin yapısına uyarlanarak optimize edilmiştir ve modelin sınıflar arası ayrım gücünü artırmayı hedeflemiştir.

4.5.3. Sınıf dengesizliğini gidermeye yönelik kayıp fonksiyonu

His analizine yönelik veri kümelerinde sıklıkla gözlenen ciddi sınıf dengesizliği problemi, modelin azınlık sınıfları öğrenmekte zorlanmasına ve bu sınıflarda düşük başarıma neden olmaktadır. Bu problemi gidermek amacıyla çalışmada sınıf dengeli odaklanmış kayıp fonksiyonu (SDOKF) kullanılmıştır. Bu fonksiyon, sınıf örnek frekanslarını dikkate alarak her örneğe ağırlık atar ve öğrenilmesi zor olan örneklere daha fazla odaklanır. Böylece, model yalnızca sık görülen örneklere değil, az görülen ve ayırt edilmesi zor hislere de duyarlı hale getirilmiştir. Bu strateji, özellikle düşük destekli sınıflarda duyarlılık ve F1 skor gibi metriklerin iyileştirilmesinde önemli katkı sağlamıştır. Buna ek olarak, sınıf dengesizliğine yalnızca kayıp fonksiyonu düzeyinde değil, veri örnekleme düzeyinde de müdahale edilmiştir. Ağırlıklı örnekleme yaklaşımıyla eğitim verilerindeki her örneğe, ait olduğu sınıfın göreli sıklığına dayalı olarak örnekleme ağırlıkları atanmış ve bu sayede modelin eğitim sırasında azınlık sınıflardan gelen örneklerle daha sık karşılaşması sağlanmıştır. Kayıp fonksiyonu ve örnekleme düzeyinde uygulanan bu stratejiler, birlikte kullanıldığında sınıf dengesizliğine karşı güçlü bir çözüm sunmuş ve modelin daha adil, dengeli ve kararlı bir sınıflama performansı göstermesine katkıda bulunmuştur.

4.5.4. Hiperparametre optimizasyonu

Modelin öğrenme performansını maksimize edebilmek için, eğitim sürecinde kritik rol oynayan hiperparametrelerin uygun değerlerde belirlenmesi gerekmektedir. Bu amaçla,

sezgisel veya manuel denemelere dayalı hiperparametre seçimi yerine, sistematik ve otomatik bir hiperparametre arama süreci benimsenmiştir.

Çalışmada hiperparametre arama süreci için, örneklemeye dayalı verimli bir arama stratejisi sunan ve modern derin öğrenme çalışmalarında yaygın olarak kullanılan Optuna çerçevesi tercih edilmiştir. Optuna, hiperparametre aramasını akıllıca yönlendiren Bayesian optimizasyon ilkeleriyle çalışmakta ve her denemenin sonucuna göre arama uzayını dinamik biçimde güncelleyebilmektedir [125].

Optuna'nın kullanılmasının başlıca gerekçeleri şu şekilde özetlenebilir:

- Arama verimliliği: Geniş dil modelleriyle çalışırken her deneme yüksek maliyetlidir. Optuna, gereksiz kombinasyonlardan kaçınarak hesaplama verimliliği sağlar.
- Hedef odaklılık: Kullanıcı tanımlı hedef metriklere göre optimizasyon yapabilir.
- Dinamik arama uzayı: Başarısız denemeleri erken durdurma ve başarılı bölgelere odaklanma özellikleriyle zamandan ve kaynaktan tasarruf sağlar.

Bu çalışmada, Optuna ile yürütülen hiperparametre optimizasyon sürecinde, doğrulama kümesindeki ağırlıklı F1 skoru hedef metrik olarak belirlenmiştir. Bu tercih, modelin yalnızca genel doğruluk değerine değil, aynı zamanda sınıflar arası dengesizliğe olan duyarlılığına da odaklanmasını sağlamıştır. Böylece, azınlık sınıflarda başarı oranı düşük kalmaksızın dengeli bir genel performans elde edilmesi amaçlanmıştır.

5. DENEYSEL ÇALIŞMA VE ELDE EDİLEN SONUÇLAR

Bu bölümde, önerilen yöntemlerin deneysel olarak değerlendirilmesi amacıyla gerçekleştirilen çalışmalar ve bu çalışmalar sonucunda elde edilen bulgular sunulmaktadır. Deneysel süreçte başarı oranları, F1 skorları ve sınıf bazlı performanslar üzerinden karşılaştırmalı analizler gerçekleştirilmiştir.

5.1. Deneysel Kurulum

Bu bölümde çalışmada gerçekleştirilen tüm deneylerin yürütüldüğü teknik alt yapı, yazılım ortamı ve deneysel ayarlara, ayrıntılı olarak yer verilmiştir. Geliştirilen modellerin sıfır atış, az atış ve ince ayar senaryolarında karşılaştırılabilir sonuçlar üretmesini sağlamak amacıyla tüm deneyler sabit bir sistemsel yapı ve tutarlı hiperparametre ayarları altında gerçekleştirilmiştir.

5.1.1. Donanım ve yazılım altyapısı

Gerçekleştirilen tüm deneyler Google Colab Pro ortamında, NVIDIA A100 GPU (40 GB) donanımı kullanılarak yürütülmüştür. Geliştirme dili olarak Python 3.10 tercih edilmiş ve Çizelge 5.1'de verilen temel kütüphaneler kullanılmıştır.

Çizelge 5.1. Deneylerde kullanılan kütüphaneler

Kütüphane	Açıklama
numpy	Sayısal işlemler, tensör ve matris hesaplamaları
pandas	Veri çerçevesi işlemleri ve ön işleme
datasets	HuggingFace veri kümesi yükleme ve yönetimi
transformers	LLaMA 3 modeli ve tokenizer arayüzü
accelerate	Hızlı model eğitimi için donanım soyutlama
bitsandbytes	4-bit quantization ve belleği verimli kullanma
peft	LoRA tabanlı parametre verimli ince ayar için özel bir modül
wandb	Deney takibi, metriklerin izlenmesi ve görselleştirme

Çizelge 5.1. (devam) Deneylerde kullanılan kütüphaneler

sentencepiece	Alt birim tabanlı tokenizer desteği
matplotlib / seaborn	Grafik ve metrik görselleştirmeleri
scikit-learn	Değerlendirme metrikleri
torch / torch.nn	Derin öğrenme modelleri ve özel loss fonksiyonları
huggingface_hub	Model/dataset erişimi ve paylaşımı
google.colab.drive	Google Drive entegrasyonu

5.1.2. Kullanılan veri setleri

Deneysel değerlendirmelerde MELD, DailyDialog ve IEMOCAP olmak üzere üç farklı çok sınıflı his analizi veri seti kullanılmıştır. Veri kümeleri resmi bölümleme oranlarına uygun olarak eğitim, doğrulama ve test alt kümelerine bölünmüştür. Çizelge 4.1, bu alt kümelerin boyutlarını göstermektedir.

5.1.3. Model ve tokenizer

Bu çalışmada, his analizi görevlerinde temel dil modeli olarak Meta AI tarafından geliştirilen LLaMA 3 Instruct mimarisi kullanılmıştır. LLaMA 3, geniş kapsamlı dil anlayışı, güçlü bağlam modelleme kapasitesi ve açık kaynak erişimi sebebiyle yapılan çalışmada tercih edilmiştir. Modelin metin girişlerini işleyebilmesi için LLaMA 3 ile tam uyumlu tokenizer, doğrudan AutoTokenizer sınıfı kullanılarak yüklenmiştir. Bu tokenizer, alt birim düzeyinde çalışarak karmaşık veya seyrek kullanılan ifadeleri anlamlı bileşenlerine ayırdığından modelin doğru yorumlama yeteneğini artırmaktadır. Özellikle diyalog içeren metinlerde sık karşılaşılan kesikli, belirsiz ya da biçimsel olarak düzensiz ifadelerin işlenmesinde bu özellik avantaj sağlamaktadır. Tokenleştirme sürecinde metinler belirli bir uzunluk sınırı çerçevesinde kırpılarak modellenmiş ve her bir örnek için otomatik hizalama mekanizması devreye alınmıştır. Bu hizalama işlemi sırasında modelin özel durdurucu simgesi aynı zamanda doldurma sembolü olarak atanmıştır.

5.1.4. Değerlendirme metrikleri

His analizi gibi çok sınıflı sınıflandırma problemlerinde modelin başarımı, en doğru şekilde karışıklık matrisi ve bu matris üzerinden türetilen çeşitli istatistiksel ölçütlerle değerlendirilmektedir. Bu matriste her bir duygu etiketi sırayla pozitif sınıf olarak ele alınmakta, diğer tüm etiketler ise negatif olarak değerlendirilerek her bir sınıf için ayrı bir değerlendirme yapılmaktadır. Örneğin, "mutluluk" sınıfına odaklanıldığında, bu sınıfa ait veriler pozitif, diğer tüm sınıflar (örneğin "üzüntü", "öfke" vb.) negatif olarak kabul edilir. Bu yaklaşım doğrultusunda, karışıklık matrisi dört temel bileşene ayrılmaktadır [123]:

- Doğru Pozitif : Hem gerçek hem de tahmin edilen sınıfın pozitif olduğu örnekler.
- Doğru Negatif : Hem gerçek hem de tahmin edilen sınıfın negatif olduğu örnekler.
- Yanlış Pozitif : Gerçekte negatif olan, model tarafından pozitif olarak tahmin edilen örnekler.
- Yanlış Negatif : Gerçekte pozitif olan, model tarafından negatif olarak tahmin edilen örnekler.

Karışıklık matrisinden türetilen kesinlik, duyarlılık, F1 skoru ve doğruluk metrikleri modelin çıktılarının doğruluğunu ve etkililiğini nicel olarak değerlendirmeye olanak sağlamaktadır. Literatüre bakıldığında modelin başarımını ölçmek için en sık kullanılan metriğin F1 skor ve olduğu, özellikle dengesiz veri kümelerinde örnek sayısıyla orantılı olarak hesaplanan ağırlıklı F1 skor tercih edildiği görülmüştür [51].

Kesinlik (Precision): Modelin pozitif olarak tahmin ettiği örnekler içerisinden gerçekten pozitif olanların oranını ifade eder. Düşük kesinlik, çok sayıda negatif örneğin yanlışlıkla pozitif olarak sınıflandırıldığını gösterir.

$$\text{Kesinlik} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Pozitif}} \tag{5.1}$$

Duyarlılık (Recall): Mevcut pozitif örnekler içerisinden kaç tanesinin model tarafından doğru bir şekilde tespit edildiğini gösterir. Düşük duyarlılık, çok sayıda pozitif örneğin gözden kaçtığını ifade eder.

$$\text{Duyarlılık} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Negatif}} \tag{5.2}$$

F1 Skoru: Kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. Bu metrik, özellikle sınıf dağılımının dengesiz olduğu durumlarda performans ölçümünde daha sağlam sonuçlar sunar.

$$\text{F1 Skor} = \frac{2 \text{ x Kesinlik x Duyarlılık}}{\text{Kesinlik + Duyarlılık}} \tag{5.3}$$

Doğruluk (Accuracy): Modelin tüm örnekler üzerindeki genel doğruluğunu ölçer. Duygu tespiti bağlamında, kesinlik tekrarlanabilirliği bu metrik ise ise doğruluğu temsil etmektedir.

$$\text{Doğruluk} = \frac{\text{Doğru Pozitif + Doğru Negatif}}{\text{Doğru Pozitif + Doğru Negatif + Yanlış Pozitif + Yanlış Negatif}} \tag{5.4}$$

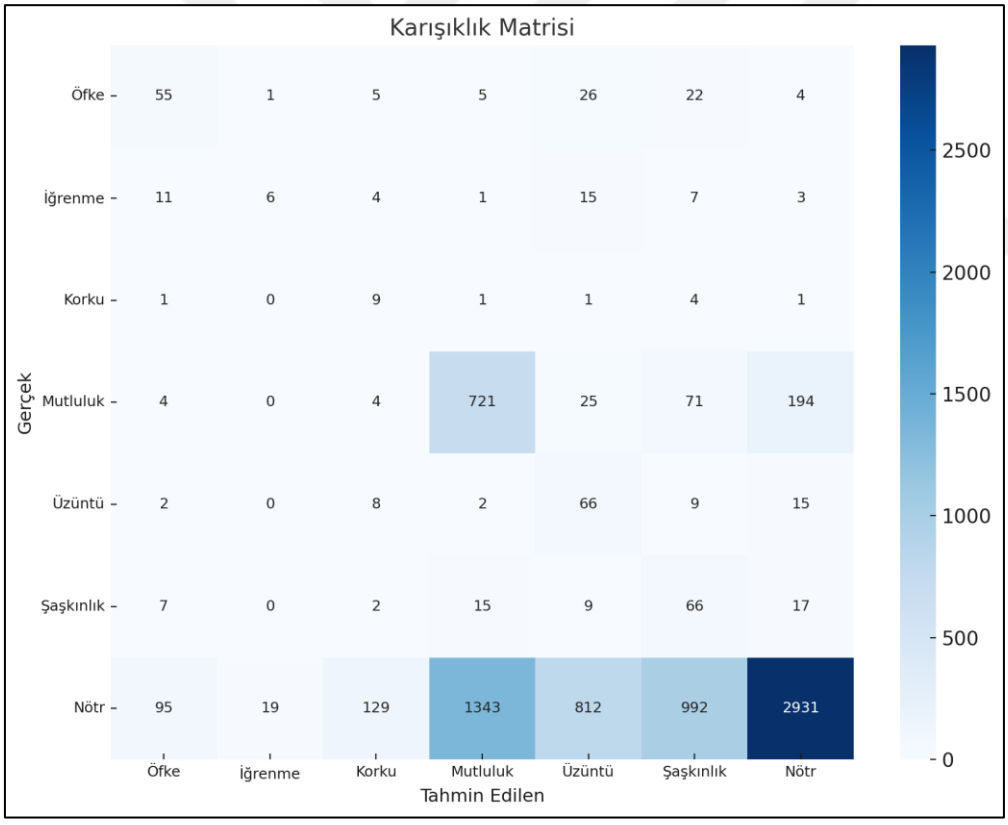
Bu dört metrik, metin tabanlı his analizi alanında en sık kullanılan kriterler arasında yer almaktadır.

5.2. Sıfır Atış Yöntemi Sonuçları

Bu bölümde, sıfır atış yaklaşımı ile elde edilen model performansı üç farklı veri kümesi için ayrı ayrı raporlanmakta ve elde edilen sonuçlar hem sınıf bazında hem de genel değerlendirme metrikleri çerçevesinde sunulmaktadır. Modelin yalnızca önceden tanımlı his etiketlerinden birini üretmesi hedeflendiğinden verilen yanıtlar normalize edilerek kontrol edilmiştir. Üretilen yanıt geçerli bir etiketle eşleşmiyorsa bilinmeyen (unknown) olarak işaretlenmiştir. Kullanılan logit işleyici sayesinde üç veri kümesi için de "bilinmeyen" etiketiyle işaretlenen bir yanıt oluşmamıştır. Sıfır atış yönteminde DailyDialog, MELD ve IEMOCAP veri kümeleri için sırasıyla %57,59; %34,67; %42,69 ağırlıklı F1 skorları elde edilmiştir. Veri kümeleri için alınan sınıf bazında sonuçlara Çizelge 5.2, 5.3 ve 5.4'te, genel değerlendirme sonuçlarına ise Çizelge 5.5'te yer verilmiştir.

Çizelge 5.2. DailyDialog veri seti sıfır atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,31	0,47	0,38	118
iğrenme	0,23	0,13	0,16	47
korku	0,06	0,53	0,10	17
mutluluk	0,35	0,71	0,46	1019
üzüntü	0,07	0,65	0,12	102
şaşkınlık	0,06	0,57	0,10	116
nötr	0,93	0,46	0,62	6321



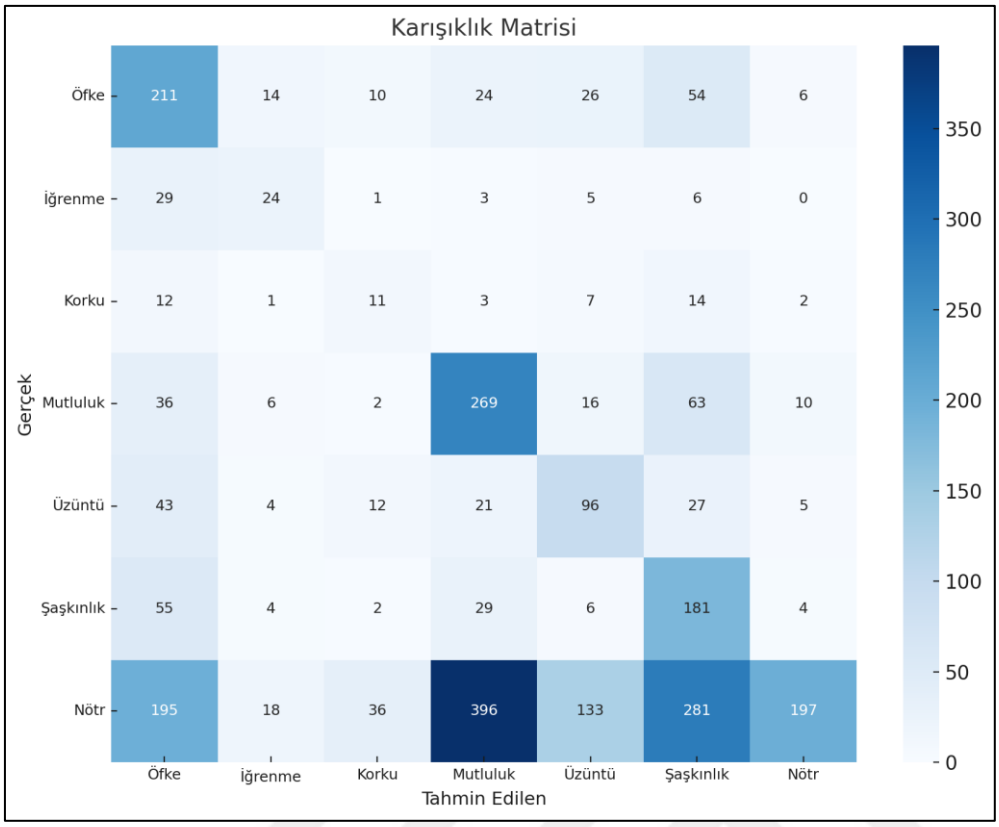
Şekil 5.1. DailyDialog veri seti sıfır atışlı öğrenme yaklaşımı için karışıklık matrisi

DailyDialog veri kümesi üzerinde sıfır atış yöntemiyle elde edilen sonuçlara göre, model özellikle “nötr” ve “mutluluk” sınıflarında diğer sınıflara göre daha başarılı bir performans göstermiştir. “Nötr” sınıfında yüksek bir kesinlik değerine ulaşılmış olsa da, duyarlılık değeri %46’da kaldığı gözlemlenmiştir. Bu da modelin bu sınıfı doğru etiketlediği

örneklerde oldukça isabetli olduğunu fakat birçok doğru örneği de gözden kaçırdığını gösterir. “Mutluluk” sınıfında ise %46 F1 skor elde edilmiş, fakat birçok pozitif örnek “nötr” olarak yanlış sınıflandırılmıştır. “Öfke”, “iğrenme”, “korku”, “üzüntü” ve “şaşkınlık” gibi düşük destekli sınıflarda hem kesinlik hem duyarlılık değerleri düşüktür. Örneğin “korku” sınıfında duyarlılık %53 gibi yüksek bir değere ulaşsa da kesinlik yalnızca %6’da kalmıştır. Bu da modelin birçok farklı örnekleri hatalı biçimde “korku” olarak etiketlediğini göstermektedir. Şekil 5.1’de verilen karışıklık matrisi incelendiğinde, modelin pek çok farklı duyguyu “nötr” ya da “mutluluk” olarak yanlış sınıflandırma eğiliminde olduğu görülmektedir. Bu durum, sınıflar arası ayrım gücünün zayıf kaldığını ve dengesiz veri dağılımının modelin ön yargı geliştirmesine neden olduğunu ortaya koymaktadır. Genel olarak, model bağlamı algılama yeteneğine sahip olsa da, düşük frekanslı sınıflar için ayırt edici özellikleri tanımlamakta zorlanmakta ve bu da sınıf bazlı F1 skorlarına doğrudan yansımaktadır.

Çizelge 5.3. MELD veri seti sıfır atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,36	0,61	0,46	345
iğrenme	0,34	0,35	0,35	68
korku	0,15	0,22	0,18	50
mutluluk	0,36	0,67	0,47	402
üzüntü	0,33	0,46	0,39	208
şaşkınlık	0,29	0,64	0,40	281
nötr	0,88	0,16	0,27	1256

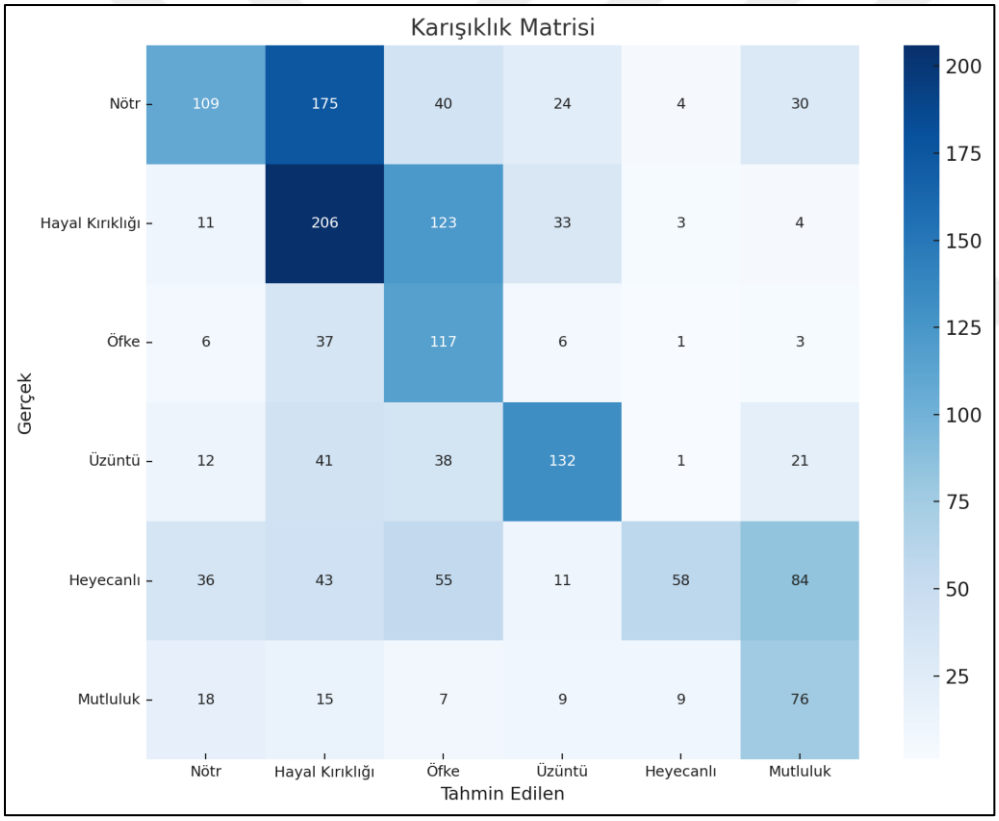


Şekil 5.2. MELD veri seti sıfır atışlı öğrenme yaklaşımı karışıklık matrisi

MELD veri kümesinde sıfır atış yöntemiyle elde edilen sınıf bazlı sonuçlar incelendiğinde, modelin özellikle “öfke” ve “mutluluk” gibi belirgin duygular üzerinde kısmen daha dengeli bir performans sergilediği görülmektedir. “Öfke” ve “mutluluk” sınıfı için %46 ve %47 F1 skoruna ulaşılması, bu sınıfın diğer duygulardan ayrıştırılmasında görece daha başarılı bir temsilin sağlandığını göstermektedir. Düşük örnek sayısına sahip “korku” ve “iğrenme” sınıflarında ise başarının oldukça sınırlı olduğu gözlemlenmiştir. Şekil 5.2’deki karışıklık matrisi incelendiğinde, “nötr” sınıfı model tarafından sıklıkla “mutluluk”, “üzüntü” ve “şaşkınlık” gibi duygularla karıştırılmakta ve bu durum yüksek sayıda yanlış pozitif etiketlemenin oluşmasına neden olmaktadır. Özellikle “nötr” sınıfında %88 kesinlik değerine rağmen duyarlılığın %16 olması modelin yalnızca çok belirgin örnekleri “nötr” olarak sınıflandırabildiğini, geri kalan pek çok örneği farklı duygularla karıştırdığını göstermektedir. Genel olarak, MELD verisi bağlamın önemli olduğu ve ironi içeren örneklerin çok olmasıyla karmaşık bir yapı sunduğundan, model bu bağlamı tek başına çözümlemede yetersiz kalmıştır.

Çizelge 5.4. IEMOCAP veri seti sıfır atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
nötr	0,57	0,29	0,38	382
hayal kırıklığı	0,40	0,54	0,46	380
öfke	0,31	0,69	0,43	170
üzüntü	0,61	0,54	0,57	245
heyecanlı	0,76	0,20	0,32	287
mutluluk	0,35	0,57	0,43	134



Şekil 5.3. IEMOCAP veri seti sıfır atışlı öğrenme yaklaşımı karışıklık matrisi

IEMOCAP veri kümesinde sıfır atış yöntemiyle elde edilen sonuçlar değerlendirildiğinde, modelin genel olarak orta düzeyde bir başarı sergilediği görülmektedir. Özellikle "üzüntü" sınıfı %57 F1 skoru ile en başarılı sınıf olurken, bu sınıf için modelin hem duyarlılığı hem de kesinliği dengeli düzeydedir. "Hayal kırıklığı" ve "öfke" sınıflarında ise sırasıyla %46 ve %43 F1 skorlarına ulaşılmış ve bu da modelin öfke ve hayal kırıklığı gibi yoğun

duyguları ayırt etmede bir ölçüde başarılı olduğunu göstermektedir. “Heyecanlı” sınıfında ise %76 kesinliğe rağmen yalnızca %20 duyarlılık elde edilmiştir. Bu değerler bu sınıfa ait gerçek örneklerin yalnızca küçük bir kısmının doğru şekilde sınıflandırılabildiğini ve modelin “heyecanlı” etiketini vermede aşırı temkinli davrandığını göstermektedir. Şekil 5.3’teki karışıklık matrisi incelendiğinde, sınıflar arası karışıklığın en çok “hayal kırıklığı” ve “öfke” etiketlerinde yaşandığı görülmektedir; her iki sınıf da sıklıkla birbirine karıştırılmıştır. Benzer şekilde, “nötr” sınıfı örneklerinin önemli bir kısmı “hayal kırıklığı” ve “öfke” olarak etiketlenmiştir. Bu durum, IEMOCAP veri kümesinde yer alan konuşmaların duygusal olarak belirsiz sınırlara sahip olması ya da farklı duyguların bir arada bulunabildiği çok katmanlı yapılar içermesiyle açıklanabilir. Sonuçlar, modelin bazı sınıfları yüksek kesinlik ile tanımasına rağmen, düşük duyarlılık nedeniyle genel başarıyı sınırladığını ve bu durumun özellikle az temsil edilen ya da semantik olarak örtüşen duygular için geçerli olduğunu göstermektedir.

Çizelge 5.5. Sıfır atışlı öğrenme yaklaşımı genel değerlendirme

Veri Seti	Ağırlıklı F1	Mikro F1	Makro F1	Örnek Sayısı
DailyDialog	0,5759	0,4979	0,2787	7740
MELD	0,3467	0,3789	0,3570	2610
IEMOCAP	0,4269	0,4368	0,4316	1598

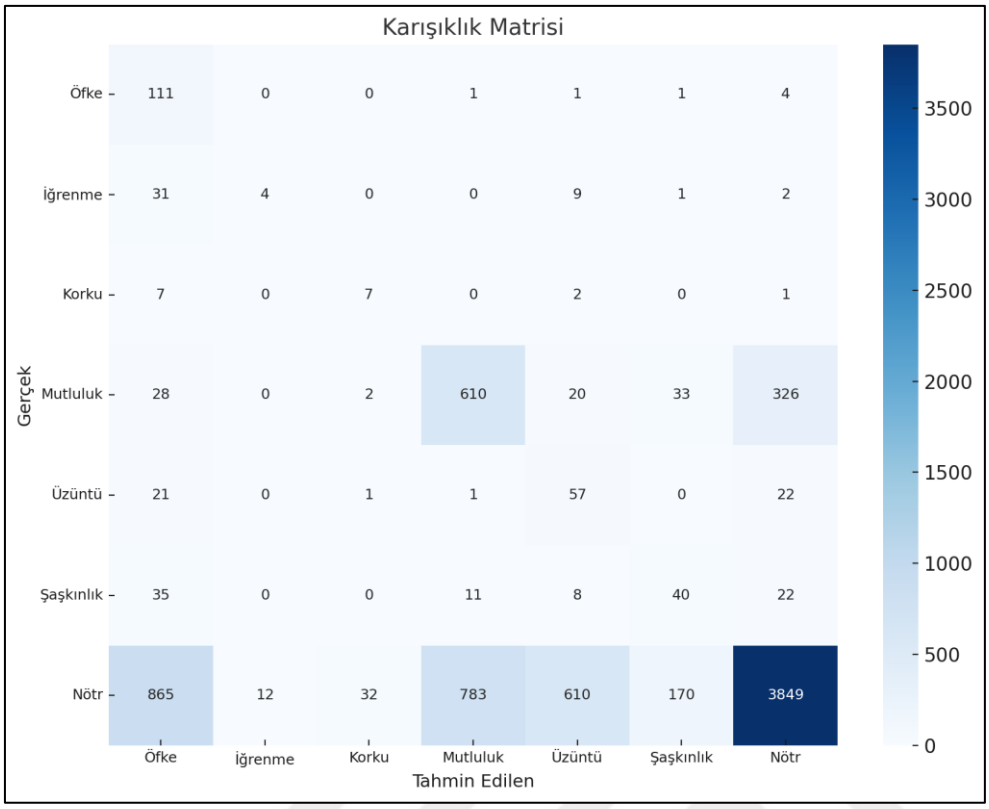
Farklı veri kümeleri üzerinde yapılan değerlendirmelerde, modelin en yüksek ağırlıklı F1 skorunu DailyDialog veri setinde elde ettiği görülmektedir. Ancak bu başarı, büyük ölçüde veri setinde baskın olarak bulunan "nötr" sınıfında yüksek doğrulukla yapılan tahminlerden kaynaklanmaktadır. Fakat aynı veri setinde makro F1 değeri oldukça düşüktür. Bu durum, modelin azınlık sınıflarda düşük başarı gösterdiğini ortaya koymaktadır. MELD veri setinde ise modelin tüm metriklerde düşük performans sergilediği görülmektedir. Bu sonuç, veri setinin bağlamsal açıdan daha karmaşık yapısından kaynaklanmakta olup benzer duyguların ayrımında yaşanan güçlükleri yansıtmaktadır. IEMOCAP veri setinde ise ağırlıklı F1, mikro F1 ve makro F1 değerlerinin birbirine yakın olduğu gözlemlenmiştir ve bu da veri setinin görece dengeli dağılımı sayesinde modelin daha istikrarlı bir performans sergileyebildiğini göstermektedir.

5.3. Az Atışlı Öğrenme Yaklaşımı Sonuçları

Bu bölümde, az atışlı öğrenme yaklaşımı ile elde edilen model performansı üç farklı veri kümesi için ayrı ayrı raporlanmakta, sonuçlar hem sınıf bazında hem de genel değerlendirme metrikleri çerçevesinde sunulmaktadır. Az atışlı öğrenme isteminde, modele görev tanımına ek olarak her sınıf için bağlamsal örnekler sunulmuş ve modelin bu örnekleri dikkate alarak hedef cümledeki duyguyu sınıflandırması beklenmiştir. Kullanılan örnekler, her sınıf için yalnızca birer temsilci içerecek şekilde, semantik ve duygusal olarak test girdisiyle en ilişkili olanlardan seçilmiştir. Yanıtların geçerliliği, önceden tanımlı his etiketleri ile eşleştirilerek denetlenmiş, eşleşmeyen yanıtlar "bilinmeyen" olarak işaretlenmiştir. Ancak çıkarım sürecine entegre edilen özel logit işleyici sayesinde model yalnızca geçerli etiketleri üretebilmiş ve bu sayede hiçbir veri kümesinde "bilinmeyen" yanıt oluşmamıştır. Az atışlı öğrenme yaklaşımı ile elde edilen ağırlıklı F1 skorları DailyDialog için %67,16; MELD için %44,12 ve IEMOCAP için %53,63 olarak hesaplanmıştır. Bu sonuçlar, uygun biçimde seçilmiş az sayıda örnekle desteklenen istemlerin, modelin his sınıflandırma performansını sıfır atış öğrenme senaryosuna kıyasla belirgin şekilde artırabildiğini göstermektedir. Veri kümelerine özgü sınıf bazlı sonuçlara Çizelge 5.6, 5.7, 5.8'da, genel karşılaştırmalara ise Çizelge 5.9'da yer verilmiştir.

Çizelge 5.6. DailyDialog veri seti az atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,10	0,94	0,18	118
iğrenme	0,25	0,09	0,13	47
korku	0,17	0,41	0,24	17
mutluluk	0,43	0,60	0,50	1019
üzüntü	0,08	0,56	0,14	102
şaşkınlık	0,16	0,34	0,22	116
nötr	0,91	0,61	0,73	6321



Şekil 5.4. DailyDialog veri seti az atışlı öğrenme yaklaşımı karışıklık matrisi

Az atışlı öğrenme senaryosunda LLaMA modeli, sınırlı sayıda örnekle yönlendirilmesine rağmen özellikle bazı duygularda yüksek duyarlılık değerlerine ulaştığı gözlemlenmiştir. Şekil 5.4'te verilen karışıklık matrisi ve Çizelge 5.6'da verilen sınıf bazlı metrikler, modelin farklı duygu sınıflarına verdiği tepkileri ayrıntılı biçimde göstermektedir.

"Öfke" sınıfı, %94 gibi oldukça yüksek bir duyarlılık değeri ile neredeyse tüm örnekleri doğru bir şekilde kapsayabilmiş ancak kesinlik oranının %10 gibi düşük bir seviyede kalması, bu sınıfa ait olmayan örneklerin de sıklıkla "öfke" olarak sınıflandırıldığını göstermektedir. Benzer şekilde, "üzüntü" ve "şaşkınlık" sınıflarında da duyarlılık yüksekken, kesinlik düşüktür. Bu durum, modelin bu sınıflara karşı aşırı duyarlı davranarak yanlış pozitifler ürettiğini gösternektedir.

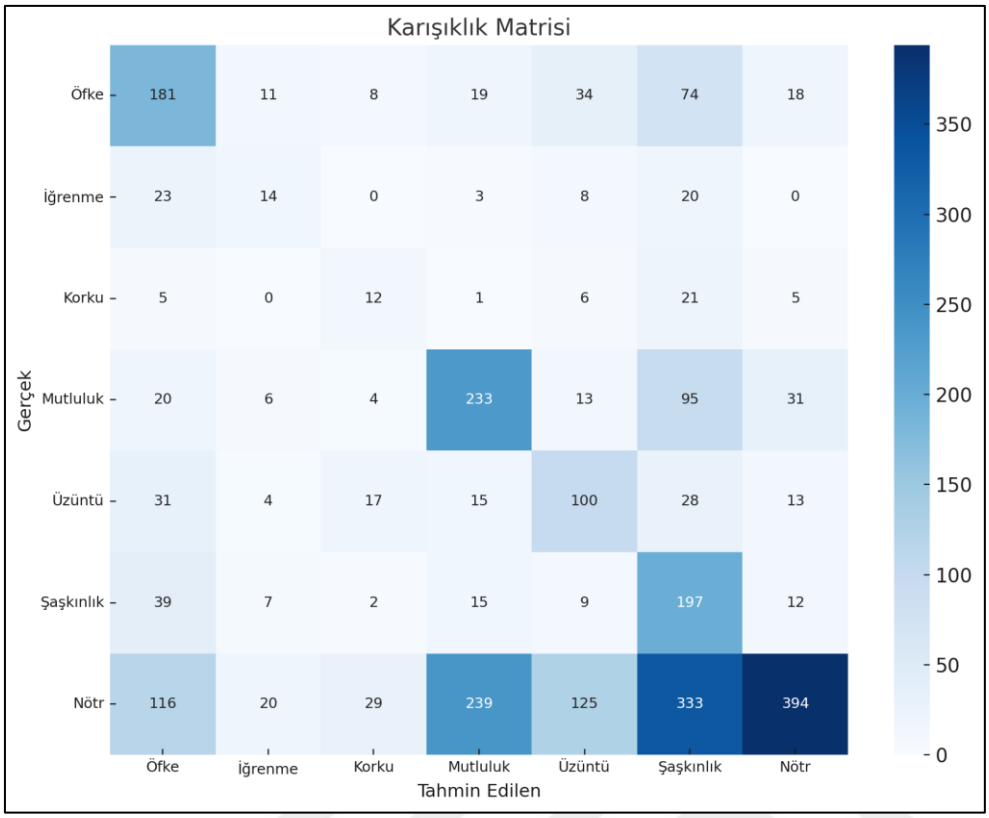
DailyDialog veri setinde en düşük performans ise "iğrenme" sınıfında gözlemlenmiştir. %9 duyarlılık ve %25 kesinlik ile bu sınıfta hem pozitif örnekler yeterince tanınamamış hem de sınıfa ait olmayan örneklerle karıştırılmıştır. "İğrenme" gibi düşük frekanslı, bağlama duyarlı ve dilsel olarak örtük duyguların zorlayıcı olduğu literatürde de vurgulanmaktadır [126, 127]. Modelin genellikle "nötr" sınıfına yönelme eğiliminde olduğu

gözlemlenmektedir. Bu eğilim, bağlamdan yeterli duygusal belirti çıkaramayan örneklerde modelin nötr sınıfa sığınmasıyla açıklanabilir.

Bu sonuçlar, LLaMA modelinin sınırlı örnekle desteklendiğinde bazı belirgin hislerde güçlü genellemeler yapabildiğini; ancak düşük destekli, semantik olarak daha soyut ya da örtük duygularda zorlandığını göstermektedir. Modelin bu bağlamda örnek temsiliyetine ve istem mühendisliğine olan bağımlılığı, az atışlı öğrenme yaklaşımının önemli sınırlılıkları arasında yer almaktadır.

Çizelge 5.7. MELD veri seti az atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,44	0,52	0,48	345
iğrenme	0,23	0,21	0,22	68
korku	0,17	0,24	0,20	50
mutluluk	0,44	0,58	0,50	402
üzüntü	0,34	0,48	0,40	208
şaşkınlık	0,26	0,70	0,38	281
nötr	0,83	0,31	0,46	1256

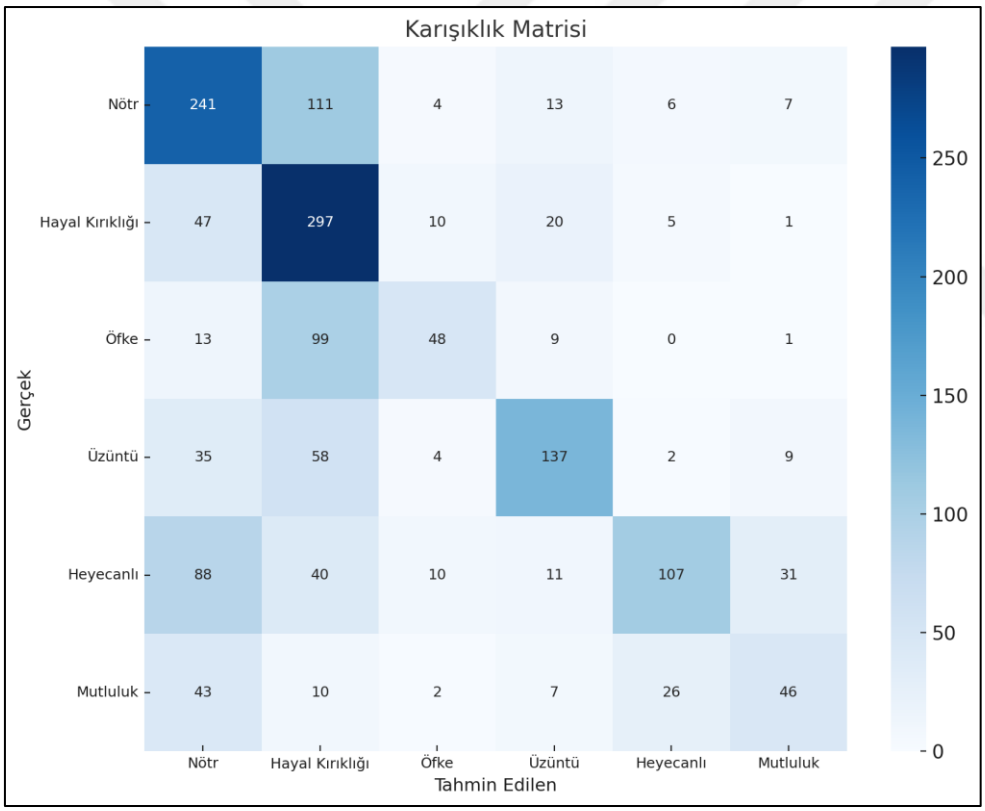


Şekil 5.5. MELD veri seti az atışlı öğrenme yaklaşımı karışıklık matrisi

MELD veri seti üzerinde yapılan az atışlı öğrenme yaklaşımında LLaMA 3 modeli, özellikle yüksek örnek frekansına sahip hisler olan “mutluluk” ve “öfke” sınıflarında görece başarılı sonuçlar elde etmiştir. Bununla birlikte, düşük destekli ve bağlamdan daha zor ayırt edilebilen “iğrenme” ve “korku” gibi duygularda belirgin bir performans düşüklüğü gözlemlenmiştir. Karışıklık matrisi analizi, modelin sıklıkla “nötr” sınıfına yöneldiğini ve özellikle “mutluluk”, “üzüntü” ve “şaşkınlık” gibi his etiketlerinin bu sınıfa hatalı biçimde etiketlendiğini ortaya koymaktadır. “Şaşkınlık” sınıfında %70 gibi yüksek bir duyarlılık elde edilmiş olsa da %26 kesinlik oranı, modelin bu sınıfa aşırı tahmin eğilimini göstermektedir. Sınıflar arası karışıklıklar, MELD veri setinin bağlam zenginliğine sahip yapısından kaynaklanan duygusal geçişlerin model tarafından sınırlı sayıda örnekle yeterince genellenemediğini göstermektedir.

Çizelge 5.8. IEMOCAP veri seti az atışlı öğrenme yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
nötr	0,52	0,63	0,57	382
hayal kırıklığı	0,48	0,78	0,60	380
öfke	0,62	0,28	0,39	170
üzüntü	0,70	0,56	0,62	245
heyecanlı	0,73	0,37	0,49	287
mutluluk	0,48	0,34	0,40	134



Şekil 5.6. IEMOCAP veri seti az atışlı öğrenme yaklaşımı karışıklık matrisi

IEMOCAP veri kümesi üzerinde yapılan az atışlı öğrenme ile his analizi deneylerinde LLaMA 3 modeli, sınırlı sayıda örnekle yönlendirildiği halde bazı duygularda dikkat çekici derecede etkili sonuçlar üretmiştir. Özellikle "hayal kırıklığı" ve "üzüntü" sınıflarında sırasıyla %60 ve %62 F1 skorlarına ulaşılması, bu duyguların bağlamsal olarak daha belirgin ve tutarlı biçimde ifade edildiği durumlarda modelin başarılı bir şekilde

genelleme yapabildiğini göstermektedir. “Nötr” sınıfı ise %57 F1 skoru ile dengeli bir performans sergilemiştir ancak modelin bu sınıfa ait örnekleri özellikle “heyecanlı” ve “hayal kırıklığı” gibi daha yoğun duygu içeren etiketlerle karıştırması, düşük duygu yoğunluklarının modellenmesinde yaşanan zorlukları göz önüne sermektedir. Öte yandan, “öfke” ve “heyecanlı” gibi yüksek uyarımlı duyguların F1 skorlarının sırasıyla %39 ve %49 seviyesinde kalması, bu sınıfların diğer duygularla semantik olarak örtüşme potansiyeli taşıması nedeniyle ayrımının daha zor olduğunu ortaya koymaktadır. Şekil 5.6’da yer alan karışıklık matrisi incelendiğinde, “öfke*”* örneklerinin önemli bir kısmının “hayal kırıklığı” ve “üzüntü” olarak etiketlendiği, “heyecanlı” örneklerinin ise sıklıkla “nötr” ve “mutluluk” ile karıştırıldığı gözlemlenmiştir. Bu durum, hisler arası sınırların bağlam içinde belirsizleştiği örneklerde modelin karar verme sürecinde zorlandığını göstermektedir. Genel olarak model, his yoğunluğu ve örnek sayısı bakımından orta seviyede temsil edilen sınıflarda daha dengeli sonuçlar verirken; olumlu, olumsuz ya da çok düşük belirginlikteki hisler arasında net ayrımlar yapmada sınırlı başarı göstermiştir.

Çizelge 5.9. Az atışlı öğrenme yaklaşımı genel değerlendirme

Veri Seti	Ağırlıklı F1	Mikro F1	Makro F1	Örnek Sayısı
DailyDialog	0,6716	0,6044	0,3060	7740
MELD	0,4412	0,4333	0,3743	2610
IEMOCAP	0,5363	0,5482	0,5113	1598

Üç farklı veri kümesi üzerinde gerçekleştirilen az atışlı öğrenme yaklaşımı deneylerinde, LLaMA 3 modelinin bağlamdan sınırlı örneklerle öğrenme yeteneği veri kümesine göre değişken performanslar sergilemiştir. En yüksek başarı DailyDialog veri setinde elde edilmiştir. Bu durum, DailyDialog’un daha yapılandırılmış ve açık ifadeli diyaloglardan oluşması nedeniyle modelin sınıf ayrımlarını daha net biçimde öğrenebildiğini göstermektedir. Ancak aynı veri setinde makro F1 değerinin düşük bir seviyede kalması, azınlık sınıflarda modelin yetersiz genelleme yaptığına işaret etmektedir. MELD veri setinde ise %44,12 ağırlıklı F1 ve %43,33 mikro F1 değerleri, MELD veri setindeki bağlamsal geçişlerin yoğun olmasının his ayrımını zorlaştırdığını göstermektedir. Bu veri setinde %37,43 makro F1 değeri, azınlık sınıfların bir miktar daha dengeli temsil edildiğini ancak yine de sınıf ayrımının yeterince güçlü olmadığını ortaya koymaktadır. IEMOCAP

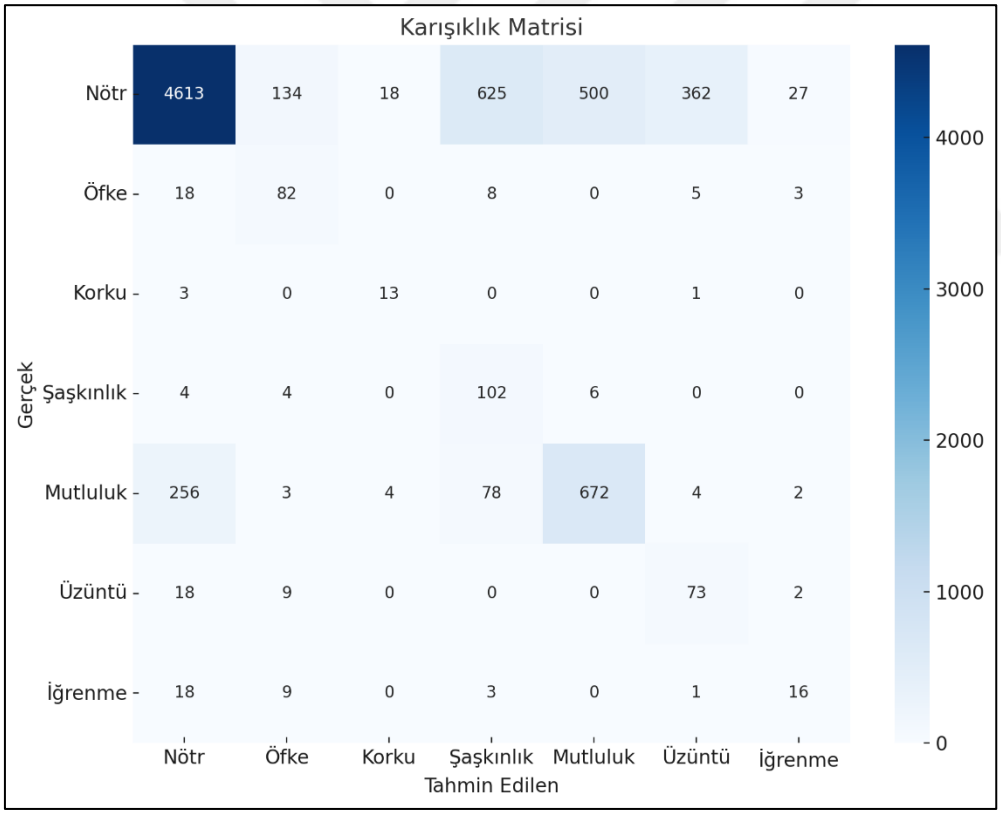
veri seti ise %53,63 ağırlıklı F1 ve %51,13 makro F1 değerleriyle, üç veri kümesi arasında sınıf dengesi açısından en istikrarlı sonuçları üretmiştir. Genel olarak değerlendirildiğinde, LLaMA 3 modeli, az atışlı öğrenme yaklaşımında güçlü bir genelleme potansiyeline sahip olmakla birlikte, veri kümesinin yapısal özellikleri ve seçilen örnekler bu performans üzerinde belirleyici bir rol oynamaktadır.

5.4. İnce Ayar Yaklaşımı Sonuçları

GDM'lerin az atış öğrenme ve sıfır atış öğrenme senaryolarında sergilediği etkileyici performansa rağmen, his sınıflandırması gibi ince ayrımlara ve semantik duyarlılığa dayalı görevlerde bu yaklaşımlar; sınıf dengesizliği, bağlamsal karmaşıklık ve düşük frekansta temsil edilen sınıflarda yetersiz genelleme gibi önemli sınırlılıklarla karşı karşıya kalmaktadır. LLaMA 3 modelinin sınırlı sayıda etiketli örnek üzerinden doğrudan eğitilmesi yoluyla gerçekleştirilen ince ayar yaklaşımı, modelin özellikle düşük frekanslı ve bağlamdan çıkarımı zor duyguları daha doğru şekilde tanıyabilmesi açısından kritik bir aşama sunmaktadır. Bu çalışmada, modelin sınıflandırma performansını artırmak amacıyla örnek ağırlıklandırmasına dayalı sınıf dengeli odaklanmış kayıp fonksiyonu, parametre verimliliği sağlayan DRU yöntemi ve düşük bellek tüketimi için 4-bit kuantizasyon stratejileri bir araya getirilmiştir. Eğitim süreci, her bir veri seti için özel olarak yapılandırılmış bağlamlı örneklerle yürütülerek modelin diyalog içinde son söylenen ifadenin duygusal bağlamını daha doğru çözümlemesi hedeflenmiştir. Bu bölümde, uygulanan ince ayar stratejisinin DailyDialog, MELD ve IEMOCAP veri kümeleri üzerindeki etkileri ayrıntılı olarak incelenmekte ve sınıf bazlı performans metrikleriyle modelin genel başarı düzeyi değerlendirilmektedir. İnce ayar yaklaşımı ile elde edilen ağırlıklı F1 skorları DailyDialog için %67,16; MELD için %44,12 ve IEMOCAP için %53,63 olarak hesaplanmıştır. Veri kümelerine özgü sınıf bazlı sonuçlara Çizelge 5.10, 5.11, 5.12'de, genel karşılaştırmalara ise Çizelge 5.13'te yer verilmiştir.

Çizelge 5.10. DailyDialog veri seti ince ayar yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,34	0,71	0,46	116
iğrenme	0,32	0,34	0,33	47
korku	0,37	0,76	0,50	17
mutluluk	0,57	0,66	0,61	1019
üzüntü	0,16	0,72	0,27	102
şaşkınlık	0,12	0,88	0,22	116
nötr	0,94	0,73	0,82	6279



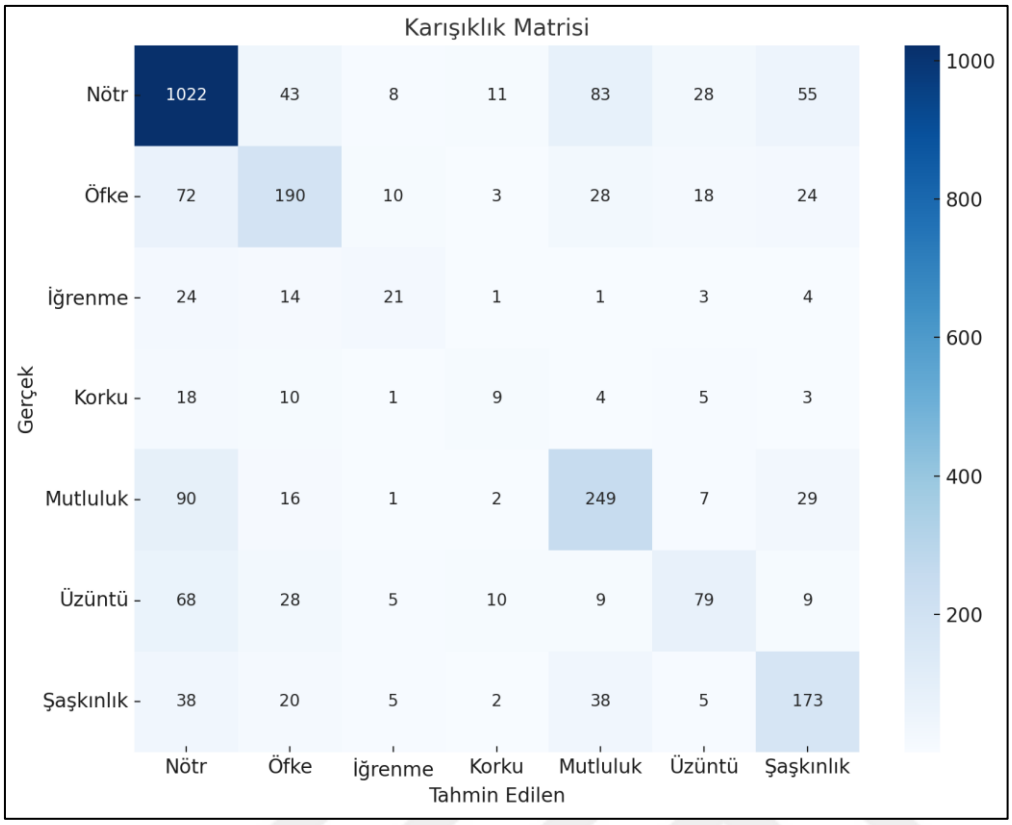
Şekil 5.7. DailyDialog veri seti ince ayar yaklaşımı karışıklık matrisi

DailyDialog veri seti üzerinde gerçekleştirilen ince ayar yaklaşımı, LLaMA 3 modelinin genel sınıflandırma performansında anlamlı bir artış sağlamıştır. Model, özellikle "nötr" sınıfında %82 F1 skoru ile elde edilen yüksek başarı, nötr ifadelerin güçlü şekilde temsil edilebildiğini ortaya koymaktadır. "Korku" sınıfında %50, "öfke" sınıfında %46 ve

“iğrenme” sınıfında %33 F1 skorlarıyla önceki az atış öğrenme sonuçlarına kıyasla önemli düzeyde artış sağlanmıştır. “Şaşkınlık” sınıfı ise %88 duyarlılık ile oldukça yüksek bir kapsama oranına ulaşmıştır fakat %12 kesinlik değeri modelin bu sınıfa aşırı sıklıkta ve hatalı tahmin yaptığını göstermektedir. Karışıklık matrisi analizi, özellikle “mutluluk” ve “üzüntü” gibi duyguların “nötr” sınıfıyla ciddi düzeyde karıştığını göstermekte, bu da modelin düşük duygusal ton barındıran olumlu ya da olumsuz ifadeleri nötr olarak yorumlama eğilimini yansıtmaktadır. %46’lık makro F1 değeri ise, modelin genel başarımının sınıflar arasında dengesiz olduğunu ve bazı duyguların öğrenilmesinde örnek yoğunluğu ile temsil gücüne bağlı farklılıkların belirleyici rol oynadığını göstermektedir. Uygulanan ince ayar yöntemi, özellikle azınlık sınıflarda belirgin performans artışları sağlayarak modelin yalnızca baskın sınıflara odaklanan değil, daha his farkında ve temsil gücü yüksek bir yapıya evrilmesine katkı sağlamıştır.

Çizelge 5.11. MELD veri seti ince ayar yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
öfke	0,59	0,55	0,57	345
iğrenme	0,41	0,31	0,35	68
korku	0,24	0,18	0,20	50
mutluluk	0,60	0,63	0,62	394
üzüntü	0,54	0,38	0,45	208
şaşkınlık	0,58	0,62	0,60	281
nötr	0,77	0,82	0,79	1250



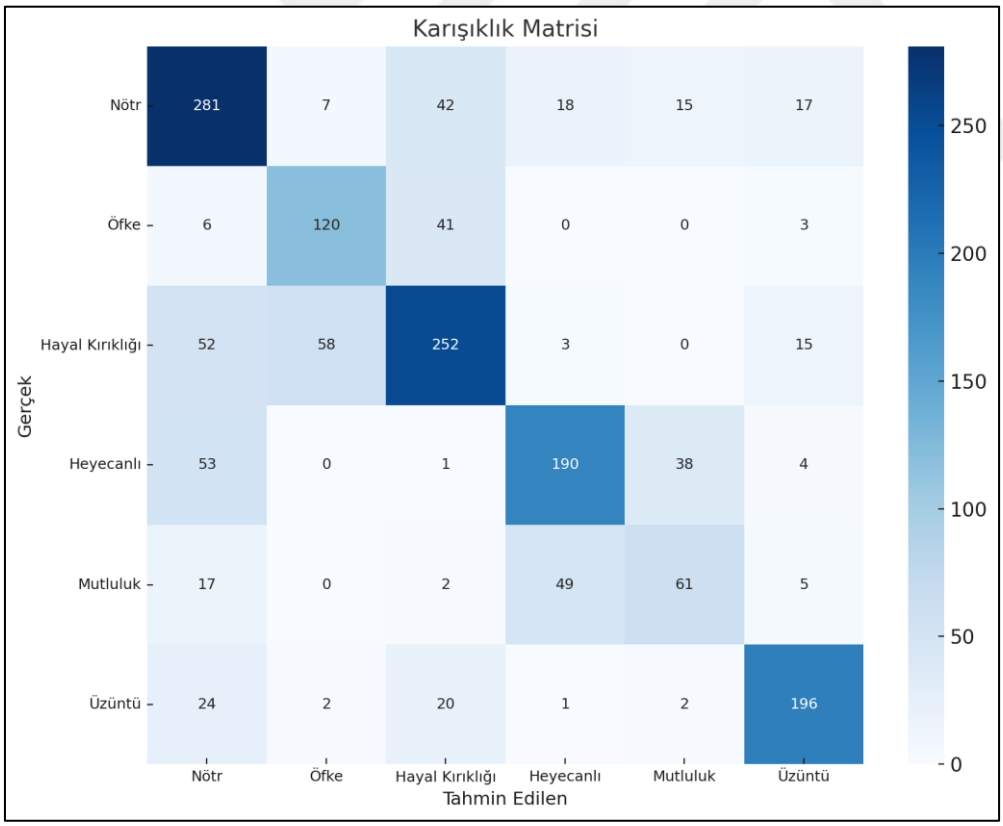
Şekil 5.8. MELD veri seti ince ayar yaklaşımı karışıklık matrisi

MELD veri seti üzerinde gerçekleştirilen ince ayar yaklaşımı sonrasında LLaMA 3 modelinin genel duygu sınıflandırma performansında belirgin bir iyileşme gözlemlenmiştir. Özellikle “mutluluk” ve “şaşkınlık” sınıflarında elde edilen yüksek skorlar, modelin olumlu ve dikkat çekici duyguların semantik özelliklerini öğrenmede başarılı olduğunu göstermektedir. “Nötr” sınıfı %79 F1 ile en yüksek performansa ulaşırken, bu durum modelin nötr ifadeleri ayırt etmede artık daha kararlı bir yapıya kavuştuğunu ortaya koymaktadır. “Öfke” sınıfında %57 F1 değerine ulaşılması, ince ayar öncesindeki az atışlı öğrenme yaklaşımı performansına kıyasla bu sınıfta kayda değer bir iyileşmeye işaret etmektedir. Öte yandan, “iğrenme” ve “korku” gibi düşük frekanslı ve bağlam bağımlılığı yüksek sınıflarda elde edilen sırasıyla %35 ve %20 F1 skorları, bu duyguların ince ayar süreciyle dahi sınırlı şekilde öğrenilebildiğini ortaya koymaktadır. Şekil 5.8’de sunulan karışıklık matrisinin analizi, modelin “üzüntü” sınıfındaki örnekleri sıklıkla “nötr”, “öfke” ve “mutluluk” gibi sınıflarla karıştırdığını göstermektedir. Bu da “üzüntü” gibi duyguların bağlam içinde diğer duygularla semantik olarak örtüşebildiğini ve sınırlarının bulanık olabileceğini düşündürmektedir. Genel olarak ince ayar süreci, yaygın sınıflarda güçlü bir performans artışı sağlarken, düşük destekli sınıflarda yapısal

sınırlılıklar devam etmiştir. Ancak genel sınıflandırma davranışında daha dengeli ve tutarlı bir dağılım elde edilmiştir.

Çizelge 5.12. IEMOCAP veri seti ince ayar yaklaşımı sınıf bazında sonuçlar

His Etiketi	Kesinlik	Duyarlılık	F1 Skor	Örnek Sayısı
nötr	0,65	0,74	0,69	380
hayal kırıklığı	0,70	0,66	0,68	380
öfke	0,64	0,71	0,67	170
üzüntü	0,82	0,80	0,81	245
heyecanlı	0,73	0,66	0,69	286
mutluluk	0,53	0,46	0,49	134



Şekil 5.9. IEMOCAP veri seti ince ayar yaklaşımı karışıklık matrisi

IEMOCAP veri seti üzerinde uygulanan ince ayar yaklaşımı sonucunda LLaMA 3 modelinin genel sınıflandırma performansında kayda değer bir iyileşme sağlanmış ve

model, tüm his sınıflarında dengeli bir başarı ortaya koymuştur. Sonuçlar modelin yalnızca baskın sınıflarda değil, tüm sınıflarda makul bir performans sergilediğini göstermektedir. Özellikle "üzüntü" sınıfında %82 F1 skoru ile ulaşılan yüksek kesinlik, bu duygunun ince ayar sürecinde belirgin bir semantik temsil kazandığını göstermektedir. Benzer şekilde, "nötr", "hayal kırıklığı", "öfke" ve "heyecanlı" sınıflarında %67 ila %69 aralığında F1 skorlarına ulaşılması, modelin farklı duygusal tonlamaları ayırt etme becerisinin güçlendiğini ortaya koymaktadır. Ancak "mutluluk" sınıfında elde edilen %49 F1 skoru, diğer sınıflara kıyasla görece düşük kalmış ve bu sınıfın özellikle "heyecanlı" ve "nötr*"* ile yüksek oranda karıştırıldığı karışıklık matrisinden açıkça gözlemlenmiştir. Aynı şekilde, "heyecanlı" ve "hayal kırıklığı" gibi yüksek uyarımlı duyguların sıklıkla birbirine karıştırılması da, duygusal yoğunluk açısından benzer sınıflar arasında modelin karar sınırlarının zaman zaman bulanıklaştığını göstermektedir. Genel tablo, modelin tüm sınıflara duyarlı hale geldiğini ve ince ayar öncesindeki dengesizliklerin büyük ölçüde giderildiğini ortaya koymaktadır.

Çizelge 5.13. İnce ayar yaklaşımı genel değerlendirme

Veri Seti	Ağirlıklı F1	Mikro F1	Makro F1	Örnek Sayısı
DailyDialog	0,7667	0,72	0,46	7696
MELD	0,66	0,67	0,51	2610
IEMOCAP	0,69	0,69	0,67	1595

Farklı yapısal özelliklere ve his dağılımına sahip üç veri kümesi üzerinde gerçekleştirilen ince ayar deneyleri, LLaMA 3 modelinin etiketli örneklerle eğitildiğinde genel sınıflandırma performansında anlamlı ve dengeli bir artış sağladığını ortaya koymaktadır. Günlük diyaloglardan oluşan ve çoğunlukla kısa, açık ifadeler içeren DailyDialog veri setinde model, %77 ağırlıklı F1 ve %72 mikro F1 skorlarına ulaşarak en yüksek genel başarıyı elde etmiştir. Ancak %46 makro F1 değeri, azınlık sınıflarda performansın sınırlı olduğunu ve bu sınıfların öğrenilmesinde veri miktarının belirleyici rol oynadığını göstermektedir. MELD veri setinde modelin %66 ağırlıklı F1 ve %67 mikro F1 skorlarıyla oldukça tutarlı bir başarı sergilediği görülmektedir. MELD veri setinin bağlam yoğunluğu düşünüldüğünde, bu sonuçlar modelin diyalog içinde duygusal geçişleri ve karşılıklı etkileşimleri daha etkili şekilde öğrenebildiğini göstermektedir. En dengeli sonuçlar ise

IEMOCAP veri setinde elde edilmiştir. IEMOCAP veri setinde %69 ağırlıklı F1, %69 mikro F1 ve %67 makro F1 değerleriyle model tüm sınıflarda istikrarlı bir performans sergilemiştir. Bu durum, IEMOCAP veri setinin his olarak daha belirgin yapılar sunması sayesinde, modelin azınlık sınıfları da dahil olmak üzere hisler arası ayrımı daha isabetli şekilde öğrenebildiğini düşündürmektedir. Genel olarak değerlendirildiğinde, ince ayar süreci modelin hem baskın hem de düşük frekanslı hisleri öğrenme kapasitesini artırmış, özellikle bağlamsal çeşitliliğe sahip veri kümelerinde sınıf bazlı dengenin iyileştirilmesine önemli katkılar sağlamıştır.

5.5. Literatürdeki Çalışmalar ile Karşılaştırma

Bu tez çalışmasında önerilen LLaMA3 8B tabanlı model, hem ağırlıklı F1 hem de mikro F1 metrikleri açısından literatürdeki güncel yöntemlerle karşılaştırıldığında, özellikle DailyDialog ve IEMOCAP veri setlerinde önemli ölçüde üstünlük sağlamıştır. Çizelge 5.14'te görüldüğü üzere, ağırlıklı F1 skoru ile LLaMA3 8B modeli DailyDialog veri setinde %76,67 ile en yakın rakibi TG-ERC skorunun üzerine çıkmıştır. IEMOCAP veri setinde ise %68,71 ağırlıklı F1 skoru ile diğer çalışmaların modellerin önünde yer almıştır. MELD veri setinde ise elde edilen %66,42'lik ağırlıklı F1, TG-ERC haricindeki tüm modellerin üzerinde yer almış ve TG-ERC ile rekabet edebilecek düzeye ulaşmıştır. Benzer şekilde, Çizelge 5.15'te sunulan mikro F1 karşılaştırmasında ise üç veri setinde de sunulan tüm çalışmalardan daha iyi bir başarı elde etmiştir.

Bu sonuçlar, yalnızca GDM'lerin genel yeteneğini değil; aynı zamanda bu çalışmada uygulanan kayıp fonksiyonları, bağlamsal örnekleme stratejileri ve çok katmanlı sınıflayıcı tasarımlarının, modelin sınıf dengesizliğine karşı direncini ve bağlam duyarlılığını artırmadaki etkisini de ortaya koymaktadır. Dolayısıyla, bu tez kapsamında geliştirilen model yapılandırması, mevcut literatürle karşılaştırıldığında hem doğruluk hem genellenebilirlik hem de azınlık sınıflar üzerindeki başarım bakımından katkı sunmaktadır.

Çizelge 5.14. Literatürdeki çalışmalar ile ağırlıklı F1 metriği karşılaştırılması

Model	DailyDialog	MELD	IEMOCAP
DialogueGCN [27]	0,5060	0,5840	0,6085
KET [48]	0,4678	0,5957	0,5816
COSMIC [29]	0,5051	0,6503	0,6343
CoG-BART [49]	0,4986	0,6481	0,6382
TODKAT [50]	0,5253	0,6133	0,6523
CFN-ESA [128]	0,5145	0,6630	0,6456
TG-ERC [51]	0,5256	0,7081	0,6541
Tez Çalışması (LLaMa 3 8B)	0,7667	0,6642	0,6871

Çizelge 5.15. Literatürdeki çalışmalar ile mikro F1 metriği karşılaştırılması

Model	DailyDialog	MELD	IEMOCAP
DialogueGCN [27]	0,5372	0,5617	0,6005
KET [48]	0,5344	0,5560	0,5988
COSMIC [29]	0,5840	0,5611	0,5889
CoG-BART [49]	0,5704	0,5880	0,6054
TODKAT [50]	0,5844	0,6475	0,6110
CFN-ESA [128]	0,5974	0,6230	0,6055
TG-ERC [51]	0,6341	0,6552	0,6120
Tez Çalışması (LLaMa 3 8B)	0,7202	0,6719	0,6896

5.6. Ablasyon Deneyleri ve Ek Gözlemler

Bu bölümde, modelin farklı bileşenlerinin his sınıflandırma performansı üzerindeki etkilerini değerlendirmek amacıyla gerçekleştirilen ablasyon testlerine yer verilmiştir. Mimaride yapılan değişikliklerin model başarımında anlamlı farklar oluşturduğu gözlemlenmiştir.

5.6.1. Çıktı kısıtlamaları ve istem tasarımının etkisi

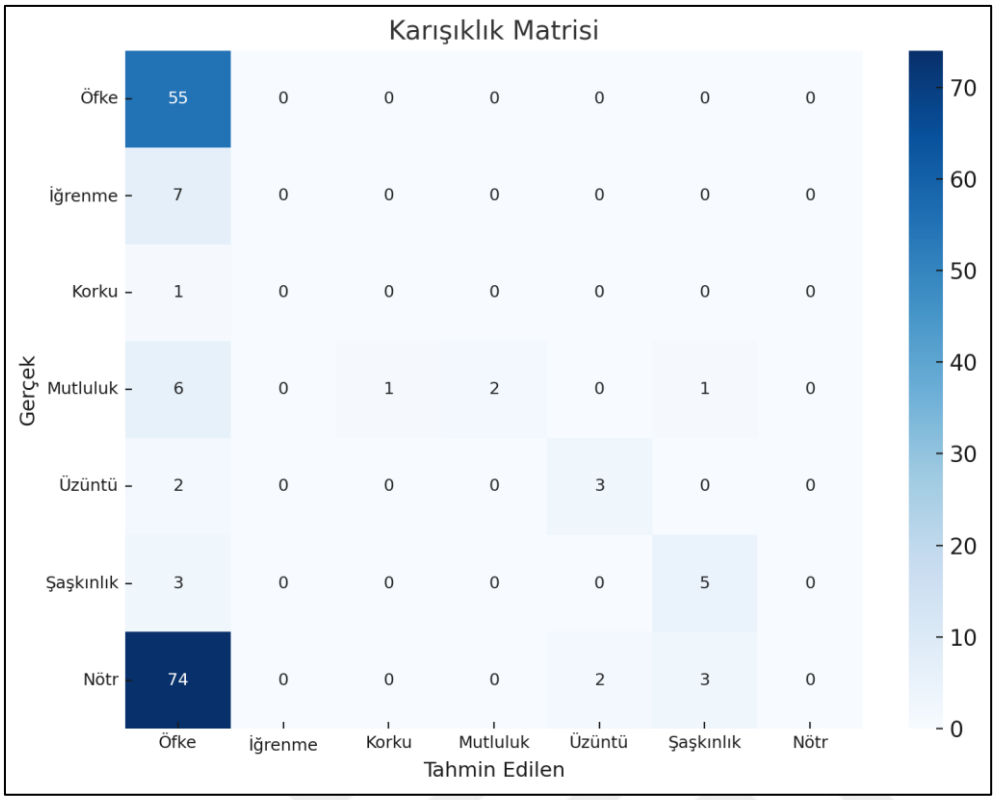
Çıktı kısıtlama mekanizmaları ile yönergeye dayalı istem bileşenlerinin bireysel katkılarını değerlendirmek amacıyla, DailyDialog veri seti kullanılarak az atışlı öğrenme yaklaşımıyla, logit işleyici modülü devre dışı bırakılmış ve istem yapısından his talimatı kaldırılmıştır. Logit işleyici ve his talimatı devre dışıyken elde edilen ayrıntılı sonuçlar Çizelge 5.16'da sunulmuştur.

Çizelge 5.16. Logit işleyici ve his talimatı kaldırılmasının etkisi

Metrik	Logit işlemciyi kaldırmanın etkisi	His talimatının kaldırmasının etkisi
Doğruluk	0,0084	0,6776
Makro F1	0,0692	0,2627
Ağırlıklı F1	0,0088	0,7143
"Bilinmeyen" sayısı	7575	3

Logit işleyicinin kaldırılmasının etkisi

Sadece geçerli his etiketlerinin üretilmesine izin veren logit işleyici devre dışı bırakıldığında, modelin performansında ciddi bir bozulma gözlemlenmiştir. Ağırlıklı F1 skoru 0,67'den 0,0088'e düşmüş ve tahminlerin %97'sinden fazlası geçersiz olmuştur. Bu mekanizma olmadan model, geçerli etiketlerle uyuşmayan uzun ve anlamsız metinler üretmiş ve neredeyse tüm örnekler "bilinmeyen" olarak sınıflandırılmıştır. Şekil 5.10'da gösterildiği üzere, logic işleyici bileşeni devre dışı bırakıldığında modelin tahminleri ciddi şekilde bozulmuştur. Özellikle, çıktıların %97'den fazlası geçerli his etiketleriyle eşleşememiş, bu da "bilinmeyen" sınıflamasında keskin bir artışa yol açmıştır.

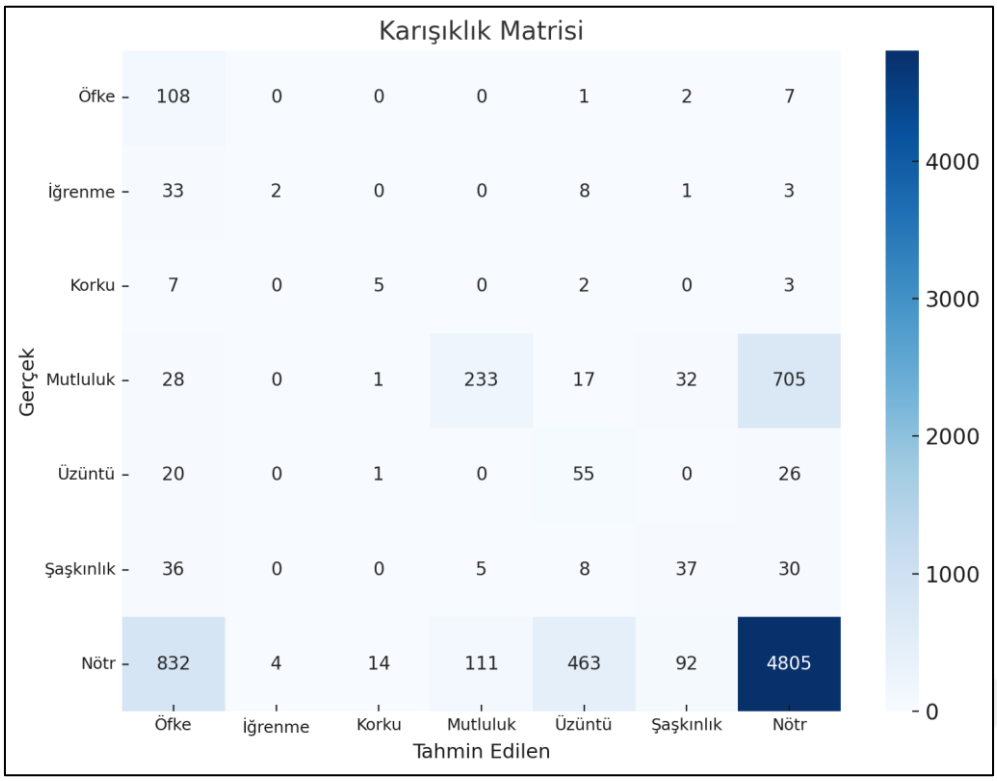


Şekil 5.10. DailyDialog veri seti az atış öğrenme yaklaşımı karışıklık matrisi (logit işlemci olmadan)

His talimatının kaldırılmasının etkisi

İstem yapısından "Emotion:" etiketi kaldırıldığında, modelin genel davranışında belirgin bir değişim gözlemlenmiştir. Ağırlıklı F1 skorundaki artış ilk bakışta olumlu bir gelişme gibi görünse de, bu artış modelin gerçek başarımını yansıtmamaktadır. Çünkü bu artış, özellikle baskın ve tahmini kolay sınıflardaki isabetli tahminlerden kaynaklanmıştır; buna karşın, seyrek veya duygusal olarak daha karmaşık sınıflarda ciddi bir belirsizlik gözlemlenmiştir. His talimatının kaldırılması tahminlerde yapısal sapmalara yol açmıştır. Nitel analizlerde modelin, his etiketlerini doğrudan belirtmek yerine bazen dolaylı ifadelerle veya açıklayıcı cümlelerle tanımladığı görülmüştür. Örneğin "öfke" yerine, etiket sistemine uymayan "Bu sinir bozucu görünüyor" gibi öznel ifadeler üretilmiştir. Sonuç olarak, "Emotion:" gibi hedef format sabitleyici öğelerin yalnızca az atışlı öğrenmede yönlendirici olmakla kalmayıp, aynı zamanda model davranışını dengeleyen kritik bir bileşen olduğu anlaşılmaktadır. Bu tür sinyallerin yokluğu, modelin tahminlerini zayıflatmakta ve özellikle kısa ya da bağlamsal olarak belirsiz ifadelerde güvenilirliğini azaltmaktadır. Bu nedenle bazı metriklerdeki artışa rağmen, görev bazlı genel

performansın düştüğü gözlemlenmiştir. Şekil 5.11'de görülebileceği üzere, talimat etiketinin kaldırılması özellikle belirsiz sınıflarda daha geniş bir tahmin dağılımına neden olmuştur.



Şekil 5.11. DailyDialog veri seti az atış öğrenme yaklaşımı karışıklık matrisi (his talimatı olmadan)

5.6.2. Kayıp fonksiyonu, örnekleme stratejisi ve sınıflayıcı derinliğinin his analizi başarımına etkileri

Bu bölümde, his analizi görevinde kullanılan modelin yapısal bileşenlerinin çıktı üzerindeki etkisini analiz etmek amacıyla gerçekleştirilen deneysel çalışmalar sunulmaktadır. Deneyler, MELD veri seti üzerinde yürütülmüş olup, özellikle sınıflandırıcı mimarinin derinliği, kullanılan kayıp fonksiyonu ve örnekleme stratejisinin model başarımına olan katkısı izole edilerek değerlendirilmiştir. Her bir deneyde yalnızca belirli bir bileşen değiştirilmiş; geri kalan tüm eğitim koşulları sabit tutularak söz konusu bileşenin çıktıya olan doğrudan etkisi ortaya konmuştur. Bu bağlamda, ilk olarak sınıflandırıcı yapısının farklı derinlik ve aktivasyon fonksiyonları ile nasıl performans değişimi yarattığı incelenmiş, ardından odaklanmalı kayıp fonksiyonu ve sınıf dengeli odaklanmalı kayıp fonksiyonu gibi farklı kayıp fonksiyonlarının sınıf dengesizliği

sorununa karşı ne ölçüde etkili olduğu değerlendirilmiştir. Son olarak, ağırlıklı rastgele örnekleyici kullanımının azınlık sınıflar üzerindeki etkisi irdelenmiş ve tüm bileşenlerin birlikte optimize edildiği yapı en iyi konfigürasyon seçilmiştir.

Çizelge 5.17. MELD veri seti üzerinde ince ayar yaklaşımı için ablasyon deneyleri

Deney Nu.	Kayıp fonksiyonu	Ağırlıklı rastgele örnekleyici	Sınıflandırıcı	Ağırlıklı F1	Mikro F1	Makro F1
1	Odaklı kayıp	Var	Lineer	0,662	0,67	0,498
2	Sınıf dengelemeli odaklı yakıp	Var	Lineer	0,631	0,60	0,48
3	Sınıf dengelemeli odaklı yakıp	Yok	2-Katman ÇKA + ReLU	0,664	0,67	0,511
4	Odaklı kayıp	Var	2-Katman ÇKA + ReLU	0,6603	0,68	0,5086
5	Odaklı kayıp	Yok	2-Katman ÇKA + ReLU	0,678	0,691	0,532

Çizelge 5.18. IEMOCAP veri seti üzerinde ince ayar yaklaşımı için ablasyon deneyleri

Deney Nu.	Kayıp fonksiyonu	Ağırlıklı rastgele örnekleyici	Sınıflandırıcı	Ağırlıklı F1	Mikro F1	Makro F1
1	Sınıf dengelemeli odaklı yakıp	Yok	2-Katman ÇKA + ReLU	0,688	0,689	0,672
2	Odaklı kayıp	Yok	2-Katman ÇKA + ReLU	0,682	0,684	0,665

His analizi modelinin başarımı üzerinde etkili olan temel bileşenlerden biri, kullanılan kayıp fonksiyonu ve bu fonksiyonun sınıf dengesizliklerine karşı duyarlılığıdır. Bu çalışma kapsamında hem odaklanmalı kayıp fonksiyonu hem de sınıf frekansına dayalı ağırlıklandırma içeren sınıf ağırlıklı odaklanmalı kayıp fonksiyonu fonksiyonları karşılaştırılmış, ayrıca ağırlıklı rastgele örnekleyici stratejisinin ve sınıflandırıcı katman derinliğinin etkileri de sistematik olarak incelenmiştir. MELD veri setinde gerçekleştirilen ablasyon deneylerinde en yüksek ağırlıklı F1 skoru deney 5 ile elde edilmiştir. Ancak bu yapılandırmada validasyon kaybı eğitim süreci boyunca hızla artmış ve aşırı öğrenme belirtisi göstermiştir. Bu durum, doğruluk oranı ve F1 skorlarının yüksek olmasına rağmen, modelin uzun vadeli başarımı açısından risk teşkil etmektedir.

Aynı sınıflandırıcı mimari, sınıf ağırlıklı odaklanmalı kayıp fonksiyonu ile yeniden eğitildiğinde ise, MELD veri seti üzerinde benzer metriklerle çalıştığı ve validasyon kaybının eğitim süreci boyunca çok daha istikrarlı bir seyir izlediği görülmüştür. Ayrıca, IEMOCAP veri setinde yapılan deneyler sınıf ağırlıklı odaklanmalı kayıp fonksiyonu tutarlılığını daha da belirginleştirmiştir. Sınıf ağırlıklı odaklanmalı kayıp fonksiyonu ile eğitilen model IEMOCAP veri setinde biraz daha yüksek skorlar elde edilmiştir. Ayrıca sınıf ağırlıklı odaklanmalı kayıp fonksiyonu ile eğitilen yapı, test setinde ağırlıklı F1 %68,8 gibi rekabetçi bir performansa ulaşmış ve validasyon kaybını eğitim boyunca birin altında tutarak daha dengeli ve güvenilir bir öğrenme eğrisi sergilemiştir.

Ayrıca ağırlıklı rastgele örnekleme stratejisinin hem MELD hem de IEMOCAP veri setlerinde sınıflandırma performansına kayda değer bir katkı sağlamadığı gözlemlenmiş, örneklemeye başvurulmadan eğitim yapılmasının modelin doğal veri dağılımına daha uygun sonuçlar verdiği anlaşılmıştır. Sınıflayıcı katmanın derinleştirilmesi ise performansı iyileştirmiştir. Tüm bu değerlendirmeler ışığında, aşırı öğrenme riski düşük, validasyon eğrisi stabil ve test seti başarımı güçlü olan sınıf ağırlıklı odaklanmalı kayıp fonksiyonunun kullanıldığı 2 katmanlı ÇKA ile sınıflandırma yapan mimari tercih edilmiştir. Bu yapı yalnızca skorlar açısından değil, modelin güvenilirliği, sınıf temsili adaleti ve genellenebilirliği gibi daha bütüncül ölçütler açısından da optimal bir denge sunmaktadır.

6. SONUÇ VE ÖNERİLER

Bu çalışma, bağlama dayalı his analizi görevinde LlaMA 3 8B GDM'nin sıfır atışlı öğrenme, az atışlı öğrenme, ince ayar gibi farklı öğrenme senaryolarında nasıl performans gösterdiğini karşılaştırmalı olarak incelemiştir. Deneysel sonuçlar, önerilen yöntemlerin hem klasik hem de çağdaş yöntemlere kıyasla önemli ölçüde üstün performans sergilediğini ortaya koymuştur. Özellikle, ince ayar senaryosunda sınıf ağırlıklı odaklanmalı kayıp fonksiyonu kullanılarak 2 katmanlı ÇKA + ReLU sınıflandırıcıyla eğitilen model, farklı veri setlerinde yüksek başarı göstermiştir.

Çalışma, literatürde yaygın olarak kullanılan yöntemlerle karşılaştırıldığında özellikle DailyDialog veri setinde 0,766'lık ağırlıklı F1 ile önemli bir sıçrama sağlamış ve TG-ERC, COSMIC gibi güçlü modellerin önüne geçmiştir. Bu durum, modelin özellikle bağlamsal bilgiye duyarlı ve genel dil modelleme yeteneklerinin, geleneksel yapay sinir ağı tabanlı yapılara kıyasla daha etkili çalıştığını göstermektedir. Ayrıca MELD ve IEMOCAP veri setlerinde de rekabetçi sonuçlar alınmıştır. Bu, modelin genellenebilirliğini ortaya koymakta ve farklı ortam ve bağlamlarda da uygulanabilirliğini göstermektedir.

Araştırma, literatüre özellikle his analizi alanında GDM'lerin sınırlı veriyi nasıl etkili kullanabileceğini göstererek katkı sunmuştur. Özellikle sıfır atış ve az atış stratejilerinin başarım düzeyleri incelenerek, örnek seçim stratejilerinin his farkındalıklı ve bağlam uyumlu hale getirilmesinin model performansına doğrudan etkisi olduğu gösterilmiştir. Bu durum, gelecekte düşük kaynaklı diller için etiketsiz ya da az etiketli verilerle etkili his analizi yapılabilmesinin mümkün olabileceğini göstermektedir.

Gelecekteki çalışmalarda önerilen model mimarisi farklı dillerde veya çok dilli his analizinde test edilebilir. Veri artırma teknikleriyle daha dengeli sınıf dağılımına sahip veri setleri oluşturularak düşük frekanslı duyguların tanınabilirliği artırılabilir. Konuşmacı kimliği, diyalog akışı ve dilsel örtük sinyaller gibi özelliklerin modele entegre edilmesiyle bağlam hassasiyeti daha da geliştirilebilir. Sadece metinsel içerikle sınırlı kalmadan, ses tonu, yüz ifadeleri ve video gibi çok modlu veriler kullanılarak modelin his tanıma kabiliyeti artırılabilir.

KAYNAKLAR

1. Akçapınar, E. ve Uçan, A. (2020). Türkçe bilgisayarlı dil bilimi çalışmalarında his analizi. *Türk Dili Araştırmaları Yıllığı-Belleten*, 70, 193-210.

2. Uçan, A. (2020). *Türkçe his analizinde optimizasyon ve ön-eğitimli modellerin kullanımı*. Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

3. Pang, B., Lee, L. and Vaithyanathan, S. (2002). *Thumbs up?: Sentiment classification using machine learning techniques.* Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP '02), Philadelphia, 79-86.

4. J Wiebe, J., Wilson, T. and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3), 165-210.

5. Aman, S., Szpakowicz, S. (Editors: V. Matoušek and P. Mautner). (2007) *Text, Speech and Dialogue*. Berlin: Springer, 196-205.

6. Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.

7. Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019). *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: Human Language Technologies (NAACL-HLT 2019), Minneapolis, 4171-4186.

8. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*. San Francisco.

9. İnternet: Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. URL: https://arxiv.org/abs/1907.11692, Son Erişim Tarihi: 26.07.2025.

10. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. and Le, Q. V. (2020). *XLNet: Generalized autoregressive pretraining for language understanding.* Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, 5753-5763.

11. LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

12. Li, R., Lin, Z., Fu, P., Wang, W. and Shi, G. (2019). EmoMix: Building an emotion lexicon for compound emotion analysis. In *Computational Science – ICCS 2019.* Berlin: Springer, 353-368.

13. Toçoğlu, M. A., Çelikten, A., Aygün, İ. ve Alpkoçak, A. (2019). Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 21(63), 719-725.

14. Park, S.-H., Bae, B.-C. and Cheong, Y.-G. (2020). *Emotion recognition from text stories using an emotion embedding model*. 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, 579-583.

15. Dheeraj, K. and Ramakrishnudu, T. (2021). Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model. *Expert Systems with Applications*, 182, 115265.

16. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

17. Yang, Y., Zhou, D., He, Y. and Zhang, M. (2019). *Interpretable relevant emotion ranking with event-driven attention*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, 177-187.

18. Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A. and Cambria, E. (2019). *DialogueRNN: An attentive RNN for emotion detection in conversations*. Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, 6818-6825.

19. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

20. Li, D., Li, Y. and Wang, S. (2020). Interactive double states emotion cell model for textual dialogue emotion prediction. *Knowledge-Based Systems*, 189, 105084.

21. Kadan, A., P, D., Gangan, M. P., Abraham, S. S. and L, L. V. (2023). REDAffectiveLM: Leveraging affect enriched embedding and transformer-based neural language model for readers' emotion detection. *Knowledge and Information Systems*, 66(12), 7495-7525.

22. Batbaatar, E., Li, M. and Ryu, K. H. (2019). Semantic-emotion neural network for emotion recognition from text. *IEEE Access*, 7, 111866-111878.

23. Hu, D., Wei, L. and Huai, X. (2021). DialogueCRN: *Contextual reasoning networks for emotion recognition in conversations*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), Bangkok, 7042-7052.

24. Hu, D., Bao, Y., Wei, L., Zhou, W. and Hu, S. (2023). *Supervised adversarial contrastive learning for emotion recognition in conversations*. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023), Toronto, 10835-10852.

25. Liu, Y., Li, J., Wang, X. and Zeng, Z. (2024). EmotionIC: Emotional inertia and contagion-driven dependency modeling for emotion recognition in conversation. *Science China Information Sciences*, 67(8), 182103.

26. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation*. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 1724-1734.

27. Ghosal, D., Majumder, N., Poria, S., Chhaya, N. and Gelbukh, A. (2019). *DialogueGCN: A graph convolutional neural network for emotion recognition in conversation*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, 154-164.

28. Lu, X., Zhao, Y., Wu, Y., Tian, Y., Chen, H. and Qin, B. (2020). *An iterative emotion interaction network for emotion recognition in conversations*. Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), Barcelona, 4078-4088.

29. Ghosal, D., Majumder, N., Gelbukh, A., Mihalcea, R. and Poria, S. (2020). *COSMIC: Commonsense knowledge for emotion identification in conversations*. Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 2470-2481

30. Jiao, W., Yang, H., King, I. and Lyu, M. R. (2019). *HiGRU: Hierarchical gated recurrent units for utterance-level emotion recognition*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Minneapolis, 397-406.

31. Mohammadi, E., Amini, H. and Kosseim, L. (2019). *Neural feature extraction for contextual emotion detection*. Proceedings - Natural Language Processing in a Deep Learning World (RANLP 2019 Workshop), Varna, 785-794.

32. Akhtar, M. S., Ekbal, A. and Cambria, E. (2020). How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble. *IEEE Computational Intelligence Magazine*, 15(1), 64-75.

33. Ishiwatari, T., Yasuda, Y., Miyazaki, T. and Goto, J. (2020). *Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), Online, 7360-7370.

34. Shen, W., Wu, S., Yang, Y. and Quan, X. (2021). *Directed acyclic graph network for conversational emotion recognition*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), Bangkok, 1551-1560.

35. Zhang, D., Chen, X., Xu, S. and Xu, B. (2020). *Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer*. Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), Barcelona, 4429-4440.

36. Calò, G., Massafra, F., De Carolis, B. and Loglisci, C. (2023). *Emotion hunters at EMit: Categorical emotion detection combining BERT and ChatGPT models*. Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), Parma, 3473-3478.

37. İnternet: Huang, Y.-H., Lee, S.-R., Ma, M.-Y., Chen, Y.-H., Yu, Y.-W. and Chen, Y.-S. (2019). EmotionX-IDEA: Emotion BERT – an affectional model for conversation. URL: http://arxiv.org/abs/1908.06264. Son Erişim Tarihi: 16.07.2025

38. Li, J., Ji, D., Li, F., Zhang, M. and Liu, Y. (2020). *HiTrans: A transformer-based context- and speaker-sensitive model for emotion detection in conversations*. Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), Barcelona, 4190-4200.

39. Li, Q., Wu, C., Wang, Z. and Zheng, K. (2020). Hierarchical transformer network for utterance-level emotion recognition. *Applied Sciences*, 10(13), 4447.

40. Nugroho, K. S. and Bachtiar, F. A. (2021). *Text-based emotion recognition in Indonesian tweet using BERT*. 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI 2021), Yogyakarta, 570-574.

41. Huang, C., Trabelsi, A. and Zaïane, O. R. (2019). *ANA at SemEval-2019 Task 3: Contextual emotion detection in conversations through hierarchical LSTMs and BERT*. Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), Minneapolis, 49-53.

42. Al-Omari, H., Abdullah, M. A. and Shaikh, S. (2020). *EmoDet2: Emotion detection in English textual dialogue using BERT and BiLSTM models*. 2020 11th International Conference on Information and Communication Systems (ICICS 2020), Irbid, 226-232.

43. Khurdula, H. V., Paguthariv, A. and Soung Yoo, J. (2024). *The future of feelings: Leveraging Bi-LSTM, BERT with attention, Palm V2 & Gemini Pro for advanced text-based emotion detection*. IEEE SoutheastCon 2024, Atlanta, 275-278.

44. Widarmanti, T., Widodo, M. P., Ramadhani, D. P. and Danlami, M. (2022). *Text emotion detection: Discover the meaning behind YouTube comments using Indo RoBERTa*. 2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS 2022), Malang, 1-6.

45. Adoma, A. F., Henry, N.-M. and Chen, W. (2020). *Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition*. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP 2020), Chengdu, 117-121.

46. Zhao, Q., Xia, Y., Long, Y., Xu, G. ve Wang, J. (2025). Leveraging sensory knowledge into text-to-text transfer transformer for enhanced emotion analysis. *Information Processing and Management*, 62(1), 103876.

47. Song, X., Huang, L., Xue, H. and Hu, S. (2022). *Supervised prototypical contrastive learning for emotion recognition in conversation*. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022), Abu Dhabi, 5197-5206.

48. Zhong, P., Wang, D. and Miao, C. (2019). *Knowledge-enriched transformer for emotion detection in textual conversations*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, 165-176.

49. Li, S., Yan, H. and Qiu, X. (2022). *Contrast and generation make BART a good dialogue emotion recognizer*. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022), Virtual/Online, 11002-11010.

50. Zhu, L., Pergola, G., Gui, L., Zhou, D. and He, Y. (2021). *Topic-driven and knowledge-aware transformer for dialogue emotion detection*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), Bangkok, 1571-1582.

51. Gou, Z., Long, Y., Sun, J. and Gao, K. (2025). TG-ERC: Utilizing three generation models to handle emotion recognition in conversation tasks. *Expert Systems with Applications*, 268, 126269.

52. Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D. and Pappas, I. (2023). A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection. *Artificial Intelligence Review*, 56(12), 15129-15215.

53. Plutchik, R. (Editor). (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*. New York: Academic Press, 3-33.

54. Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M. and Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712-717.

55. Picard, R. W. (1997). Affective computing. Cambridge, MA: MIT Press, 166-192.

56. Soergel, D. (1998). WordNet: An electronic lexical database. *Language*, 76(3), 706-708.

57. Cambria, E., Speer, R., Havasi, C. and Hussain, A. (2010). *SenticNet: A publicly available semantic resource for opinion mining*. AAAI 2010 Fall Symposium on Commonsense Knowledge: Papers from the AAAI Fall Symposium, Arlington, VA, 14-18.

58. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. Advances in Neural Information Processing Systems (NeurIPS 2013), Lake Tahoe, NV, 3111-3119.

59. Kalchbrenner, N., Grefenstette, E. and Blunsom, P. (2014). *A convolutional neural network for modelling sentences*. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), Volume 1: Long Papers, Baltimore, MD, 655-665.

60. Kim, Y. (2014). *Convolutional neural networks for sentence classification*. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, 1746-1751.

61. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, 5998-6008.

62. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V. and Salakhutdinov, R. (2019). *Transformer-XL: Attentive language models beyond a fixed-length context*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 2978-2988.

63. İnternet: Rosset, C. (February, 2020). Turing-NLG: A 17-billion-parameter language model by Microsoft. Microsoft Research Blog. URL: https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/. Son Erişim Tarihi: 09.02.2025.

64. Singh, R. and Kaur, R. (2015). Sentiment analysis on social media and online review. *International Journal of Computer Applications*, 121(20), 44-48.

65. Afzal, S., Ali Khan, H., Jalil Piran, M. and Weon Lee, J. (2024). A comprehensive survey on affective computing: Challenges, trends, applications, and future directions. *IEEE Access*, 12, 96150-96168.

66. Bruna, O., Avetisyan, H. and Holub, J. (2016). Emotion models for textual emotion classification. *Journal of Physics: Conference Series*, 772, 012063.

67. Binali, H. and Potdar, V. (2012). *Emotion detection state of the art*. Proceedings of the CUBE International Information Technology Conference, Pune, India, 501-507.

68. Jin, X. and Wang, Z. (Editors: J. Tao, T. Tan and R. W. Picard) (2005). *Affective computing and intelligent interaction*. Beijing: Springer, 397-402.

69. Gunes, H. and Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, 1(1), 68-99.

70. Colby, B. N. (1989). Describing the emotions: a review of the cognitive structure of emotions, by Ortony, Clore and Collins. *Contemporary Sociology*, 18(6), 957.

71. McCrae, R. R. and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90.

72. Strapparava, C. and Mihalcea, R. (2007). *SemEval-2007 task 14: Affective text*. Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic, 70-74.

73. Mohammad, S. and Bravo-Marquez, F. (2017). *WASSA-2017 shared task on emotion intensity*. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2017), Copenhagen, Denmark, 34-49.

74. Araque, O., Frenda, S., Sprugnoli, R., Nozza, D. and Patti, V. (2023). *EMit at EVALITA 2023: Overview of the categorical emotion detection in Italian social media task*. Proceedings of the EVALITA 2023 Workshop, Parma, Italy.

75. Scherer, K. R. and Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2), 310-328.

76. Buechel, S. and Hahn, U. (2017). *EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis*. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017), Volume 2: Short Papers, Valencia, 578-585.

77. Agrawal, P. and Suri, A. (2019). *NELEC at SemEval-2019 task 3: Think twice before going deep*. Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), Minneapolis, MN, 266-271.

78. Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S. and Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4), 335-359.

79. Lu, X., Yu, Z., Guo, B., Zhang, J., Chin, A., Tian, J. and Cao, Y. (2014). *Trending words based event detection in Sina Weibo*. Proceedings of the 2014 International Conference on Big Data Science and Computing, Beijing, China, 1-6.

80. İnternet: Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G. and Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. URL: https://doi.org/10.48550/arXiv.2005.00547. Son Erişim Tarihi: 16.07.2025.

81. Li, J. and Ren, F. (2011). Creating a Chinese emotion lexicon based on corpus Ren-CECps. *Proceedings of the 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Beijing, China, 80-84.

82. Li, Y., Su, H., Shen, X., Li, W., Cao, Z. and Niu, S. (2017). *DailyDialog: A manually labelled multi-turn dialogue dataset*. Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017), Volume 1: Long Papers, Taipei, Taiwan, 986-995.

83. Arslan, M., Mubeen, M., Akram, A., Abbasi, S. F., Ali, M. S. and Tariq, M. U. (2024). *A deep features based approach using modified ResNet50 and gradient boosting for visual sentiments classification*. Proceedings of the 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 239-242.

84. Mohammad, S. (2012). *#Emotional tweets*. Proceedings of the First Joint Conference on Lexical and Computational Semantics (SemEval 2012), Montréal, Canada, 246-255.

85. Liu, V., Banea, C., and Mihalcea, R. (2017). *Grounded emotions*. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 477–483.

86. Kim, H., Kim, B., and Kim, G. (2021). *Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes*. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2227–2240.

87. Schuff, H., Barnes, J., Mohme, J., Padó, S., and Klinger, R. (2017). *Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus*. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2017), Copenhagen, Denmark, 13–23.

88. Tocoglu, M. A., and Alpkocak, A. (2018). TREMO: A dataset for emotion analysis in Turkish. *Journal of Information Science, 44*(6), 848–860.

89. İnternet: Chen, S.-Y., Hsu, C.-C., Kuo, C.-C., Huang, T.-H., and Ku, L.-W. (2018). EmotionLines: An emotion corpus of multi-party conversations. URL: https://doi.org/10.48550/arXiv.1802.08379, Son Erişim Tarihi: 13.04.2025.

90. Tocoglu, M. A., Ozturkmenoglu, O., and Alpkocak, A. (2019). Emotion analysis from Turkish tweets using deep neural networks. *IEEE Access*, 7, 183061-183069.

91. Nojavanasghari, B., Baltrušaitis, T., Hughes, C. E., and Morency, L.-P. (2016). *EmoReact: A multimodal approach and dataset for recognizing emotional responses in children*. Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 137-144.

92. Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kohli, P., and Allen, J. (2016). *A corpus and cloze evaluation for deeper understanding of commonsense stories*. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2016), San Diego, CA, USA, 839-849.

93. Saputri, M. S., Mahendra, R., and Adriani, M. (2018). *Emotion classification on Indonesian Twitter dataset*. 2018 International Conference on Asian Language Processing (IALP), Bandung, Indonesia, 90-95.

94. Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., and Mihalcea, R. (2019). *MELD: A multimodal multi-party dataset for emotion recognition in conversations*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), Florence, Italy, 527-536.

95. Kadan, A., Patil, D., Abraham, S. S., V. L., L., and Gangan, M. P. (2022). Readers' affect: Predicting and understanding readers' emotions with deep learning. *Journal of Big Data*, 9(1), 82.

96. Alm, C. O., Roth, D., and Sproat, R. (2005). *Emotions from text: Machine learning for text-based emotion prediction*. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), Vancouver, Canada, 579-586.

97. Mohammad, S. M., Zhu, X., Kiritchenko, S., and Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4), 480-499.

98. İnternet: Zahiri, S. M., and Choi, J. D. (2017). Emotion detection on TV show transcripts with sequence-based convolutional neural networks. URL: https://doi.org/10.48550/arXiv.1708.04299. Son Erişim Tarihi: 29.05.2025.

99. Kalra, V., and Aggarwal, R. (2018). *Importance of text data preprocessing & implementation in RapidMiner*. 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, Czech Republic, 71-75.

100. Uysal, A. K., and Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, 50(1), 104-112.

101. Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2010). *Part-of-speech tagging for Twitter: Annotation, features, and experiments*. Defense Technical Information Center (DTIC), Portland, Oregon, 42-47.

102. Balahur, A. (2013). *Sentiment analysis in social media texts*. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2013), Atlanta, GA, USA, 120-128.

103. Giachanou, A., Gonzalo, J., Mele, I., and Crestani, F. (Editors: J. M. Jose, C. Hauff, I. S. Altingovde, D. Song, D. Albakour, S. Watt, and J. Tait) (2017). *Advances in information retrieval*. Cham: Springer, 226-238.

104. Balazs, J. A., and Velásquez, J. D. (2016). Opinion mining and information fusion: A survey. *Information Fusion*, 27, 95-110.

105. Keya, A. J., Wadud, M. A. H., Mridha, M. F., Alatiyyah, M., and Hamid, M. A. (2022). AugFake-BERT: Handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification. *Applied Sciences*, 12(17), 8398.

106. Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). *Best practices for convolutional neural networks applied to visual document analysis*. Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), Edinburgh, Scotland, 958-963.

107. Bayer, M., Kaufhold, M.-A., Buchhold, B., Keller, M., Dallmeyer, J., and Reuter, C. (2023). Data augmentation in natural language processing: A novel text generation approach for long and short text classifiers. *International Journal of Machine Learning and Cybernetics*, 14(1), 135-150.

108. Thakkar, G., Preradović, N. M., and Tadić, M. (2024). Examining sentiment analysis for low-resource languages with data augmentation techniques. *Eng*, 5(4), 2920-2942.

109. Chen, H., Dan, L., Lu, Y., Chen, M., and Zhang, J. (2024). An improved data augmentation approach and its application in medical named entity recognition. *BMC Medical Informatics and Decision Making*, 24(1), 221.

110. Kapusta, J., Držík, D., Šteflovič, K., and Nagy, K. S. (2024). Text data augmentation techniques for word embeddings in fake news classification. *IEEE Access*, 12, 31538-31550.

111. Xiang, R., Chersoni, E., Lu, Q., Huang, C., Li, W., and Long, Y. (2021). Lexical data augmentation for sentiment analysis. *Journal of the Association for Information Science and Technology*, 72(11), 1432-1447.

112. Zhao, H., Chen, H., Ruggles, T. A., Feng, Y., Singh, D., and Yoon, H.-J. (2024). Improving text classification with large language model-based data augmentation. *Electronics*, 13(13), 2535.

113. Luo, H. (2021). Emotion detection for Spanish with data augmentation and transformer-based models. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021)*, Málaga, Spain, 35-42.

114. İnternet: Poslavskaya, E., and Korolev, A. (2023). Encoding categorical data: Is there yet anything 'hotter' than one-hot encoding. URL: https://doi.org/10.48550/arXiv.2312.16930. Son Erişim Tarihi: 22.06.2025.

115. Aslam, N., Rustam, F., Lee, E., Washington, P. B., and Ashraf, I. (2022). Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model. *IEEE Access*, 10, 39313-39324.

116. Qaiser, S., and Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25-29.

117. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., and Kalai, A. (2016). *Man is to computer programmer as woman is to homemaker? Debiasing word embeddings*. Advances in Neural Information Processing Systems (NeurIPS 2016), Barcelona, Spain, 4356-4364.

118. Nandwani, P., and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81.

119. Aka Uymaz, H., and Kumova Metin, S. (2022). Vector based sentiment and emotion analysis from text: A survey. *Engineering Applications of Artificial Intelligence*, 113, 104922.

120. Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 1532–1543.

121. Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.

122. Sarbazi-Azad, S., Akbari, A., and Khazeni, M. (2021). *ExaAEC: A new multi-label emotion classification corpus in Arabic tweets*. 2021 11th International Conference on Computer Engineering and Knowledge (ICCKE), Mashhad, Iran, 465–470.

123. Maruf, A. A., Khanam, F., Haque, M. M., Jiyad, Z. M., Mridha, M. F., and Aung, Z. (2024). Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*, 12, 18416–18450.a

124. İnternet: Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., Ma, Z. (2024). The Llama 3 Herd of Models. URL*:* https://doi.org/10.48550/arXiv.2407.21783. Son Erişim Tarihi: 03.08.2025.

125. Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). *Optuna: A next-generation hyperparameter optimization framework*. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 2623–2631.

126. Meng, T., Shou, Y., Ai, W., Yin, N., and Li, K. (2024). Deep imbalanced learning for multimodal emotion recognition in conversations. *IEEE Transactions on Artificial Intelligence*, 5(12), 6472–6487.

127. Kong, L., Yang, W., and Wei, F. (2023). Implicit emotion analysis based on improved supervised contrastive learning. *Electronics*, 12(10), 2172.

128. Li, J., Wang, X., Liu, Y., and Zeng, Z. (2024). CFN-ESA: A cross-modal fusion network with emotion-shift awareness for dialogue emotion recognition. *IEEE Transactions on Affective Computing*, 15(4), 1919–1933.

Gazili olmak ayrıcalıktır