

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**TÜRKÇE E-TİCARET YORUMLARINDA
TRANSFORMER TABANLI DUYGU ANALİZİ**

YÜKSEK LİSANS TEZİ

Muhammet Şamil İKİZOĞLU

İşletme Mühendisliği Anabilim Dalı

İşletme Mühendisliği Programı

OCAK 2026

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**TÜRKÇE E-TİCARET YORUMLARINDA
TRANSFORMER TABANLI DUYGU ANALİZİ**

YÜKSEK LİSANS TEZİ

**Muhammet Şamil İKİZOĞLU
(507211062)**

İşletme Mühendisliği Anabilim Dalı

İşletme Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Tuncay ÖZCAN

OCAK 2026

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**TRANSFORMER-BASED SENTIMENT ANALYSIS
ON TURKISH E-COMMERCE REVIEWS**

M.Sc. THESIS

**Muhammet Şamil İKİZOĞLU
(507211062)**

Department of Management Engineering

Management Engineering Programme

Thesis Advisor: Doç. Dr. Tuncay ÖZCAN

JANUARY 2026

İTÜ, Lisansüstü Eğitim Enstitüsü'nün 507211062 numaralı Yüksek Lisans Öğrencisi Muhammet Şamil İKİZOĞLU, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "TÜRKÇE E-TİCARET YORUMLARINDA TRANSFORMER TABANLI DUYGU ANALİZİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Tuncay ÖZCAN**

İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Prof. Dr. Tolga Kaya**

İstanbul Teknik Üniversitesi

Prof. Dr. Tarık Küçükdeniz

İstanbul Üniversitesi - Cerrahpaşa

Teslim Tarihi : **29 Aralık 2025**

Savunma Tarihi : **23 Ocak 2026**





Eşime ve yeni doğacak kızıma,



ÖNSÖZ

Öncelikle, tez süreci boyunca akademik bilgisi, tecrübesi ve kıymetli yönlendirmeleri ile çalışmama ve şahsıma değer kazandıran tez danışmanım Doç. Dr. Tuncay ÖZCAN'a özel olarak teşekkür ederim. Kendisi, akademik yetkinliği ve güçlü insani yönü sayesinde yoğun emek gerektiren bu tez çalışması sürecini çok başarılı bir şekilde yöneterek nitelikli bir çalışma ortaya konulmasında önemli bir paya sahiptir.

Çalışma hayatımla tez sürecini dengeli bir şekilde yürütebilmem için anlayış ve destek gösteren Mercan Optik ailesine ve şirket yöneticilerine ayrıca teşekkür ederim. Yoğun iş temposu içerisinde tezimi ilerletebilmem adına sağladıkları esneklik, güven ve teşvik benim için son derece değerli olmuştur.

Bana her zaman güç veren ve hiçbir zaman desteğini esirgemeyen aileme ve eşime en derin teşekkürlerimi sunarım. Bu süreçte bana en büyük motivasyon kaynağım olan doğacak kız çocuğumun da gözlerinden öperim. Mezuniyette görüşmek üzere güzel kızım.

Ocak 2026

Muhammet Şamil İKİZOĞLU

İÇİNDEKİLER

Sayfa

ÖNSÖZ	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
ÇİZELGE LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1. GİRİŞ	1
2. LİTERATÜR ARAŞTIRMASI	3
2.1 Sistematik Literatür Taramaları	3
2.2 Ampirik Çalışmalar	5
3. METODOLOJİ	25
3.1 Doğal Dil İşleme	25
3.2 Duygu Analizi	27
3.3 Yöntem ve Modeller	28
3.3.1 Kural tabanlı yöntemler	28
3.3.1.1 Duygu sözlük tabanlı yaklaşım	29
3.3.1.2 Bağlamsal kutuplaşma yaklaşımı	29
3.3.2 Makine öğrenimi ve derin öğrenme yöntemleri	29
3.3.2.1 Tamamlayıcı naif bayes	30
3.3.2.2 K-en yakın komşuluk	31
3.3.3 Transformer tabanlı modeller	31
3.3.3.1 BERT	32
3.3.3.2 RoBERTa	33
3.3.3.3 ELECTRA	34
3.3.3.4 DeBERTa	35
3.3.3.5 BERTurk	36
3.3.3.6 TurkishBERTweet	36
3.3.3.7 ModernBERT	37
3.3.4 Büyük dil modelleri ve üretken yapay zeka	38
3.4 Model Eğitimleri	41
3.4.1 İnce ayar mimarisi	41
3.4.1.1 Tam ince ayar	42
3.4.1.2 Parametre verimli ince ayar	43
3.4.1.3 Talimat temelli ince ayar	44
3.4.2 Hiperparametre optimizasyonu	45
3.4.2.1 Klasik optimizasyon	46
3.4.2.2 Bayes optimizasyon	48
3.5 Performans Metrikleri	50
3.5.1 Sınıf bazlı metrikler	50

3.5.1.1 Doğruluk.....	51
3.5.1.2 Kesinlik	52
3.5.1.3 Duyarlılık	52
3.5.1.4 Özgüllük	52
3.5.1.5 F1 Skoru	53
3.5.2 Genel metrikler.....	53
3.5.2.1 Mikro ortalama	53
3.5.2.2 Makro ortalama	54
3.5.2.3 Ağırlıklı ortalama	54
4. UYGULAMA.....	57
4.1 Veri Madenciliği Süreci	57
4.2 Veri Setlerine Genel Bakış	59
4.2.1 Ürün detay	60
4.2.2 Ürün yorumları.....	62
4.2.3 Ürün soru ve cevapları	63
4.3 Veri Etiketleme.....	64
4.4 Ön İşleme.....	68
4.4.1 Normalizasyon	69
4.4.2 Küçük harflere dönüştürme	70
4.4.3 Gürültü azaltma	70
4.4.4 Türkçe karakter düzenlemeleri.....	70
4.4.5 Emoji dönüştürme	70
4.4.6 Harf tekrarlarını azaltma	71
4.4.7 Noktalama, işaret ve boşluk düzenlemeleri.....	71
4.4.8 Yorum dil tespiti.....	71
4.5 Model Eğitimleri	71
4.5.1 Nihai veri seti	73
4.5.2 Eğitim, doğrulama ve test ayrımı	78
4.5.3 Temel model araştırmaları	80
4.5.4 Varsayılan parametreler ile eğitimler.....	83
4.5.5 Optuna parametre optimizasyonu ile eğitimler	84
4.6 Sonuçların Değerlendirilmesi	88
4.7 Modellerin Yayınlanması	96
5. SONUÇLAR	99
6. GELECEĞE İLİŞKİN ÖNERİLER.....	101
KAYNAKLAR.....	103
ÖZGEÇMİŞ.....	113

KISALTMALAR

ALBERT	: Hafifletilmiş BERT Modeli (ALBERT)
ALBERT-TR	: Türkçe ALBERT Modeli (ALBERT-Tr)
APSO-UKSB	: APSO Optimizasyonlu UKSB (APSO-LSTM)
BDM	: Büyük Dil Modelleri (LLM)
BERT	: Çift Yönlü Transformer Tabanlı Kodlayıcı Temsilleri (BERT)
BERT-M	: Çok Dilli BERT (BERT-M/ mBERT)
BERTurk	: Türkçe BERT (BERTurk)
BERTweet	: Tweet Tabanlı BERT Model (BERTweet)
BNB	: Bernoulli Naive Bayes (BNB)
BO	: Bayes Optimizasyonu (BO)
BOOL	: Boole Özellikleri (BOOL)
BO-DVM	: Boole özellikli DVM (BOOL-SVM)
CIE	: Karakter Tamsayı Gömme (CIE)
ÇK-ESA	: Çok Kanallı CNN (MCNN)
ÇO	: Çoğunluk Oyu (MV)
COE	: Karakter Tek-Sıcak Kodlama Gömme (COE)
ÇY-UKSB	: Çift Yönlü Uzun Kısa Süreli Bellek (BiLSTM)
ÇY-UKSB-ESA	: ÇY-UKSB–ESA Hibrit (BiLSTM-CNN)
DA	: Duygu Analizi (SA)
DDİ	: Doğal Dil İşleme (NLP)
DDİY	: Dilbilimsel Doğal Dil İşleme Yöntemleri (Ling-NLP)
DDVM	: Doğrusal Destek Vektör Makinesi (LSVC)
DeBERTa	: Ayrıştırılmış Dikkat Mekanizmalı ve Çözümleme Geliştirmeli BERT (DEBERTA)
DistilBERT	: Sıkıştırılmış BERT Modeli (DistilBERT)
DR	: Doğrusal Regresyon (LR)
DTT	: Değiştirilen Token Tespiti (RTD)
DVM	: Destek Vektör Makineleri (SVM)
ELECTRA	: Verimli Yerine-Koyma Tabanlı Kodlayıcı (ELECTRA)
ELECTRA-TR	: Türkçe ELECTRA Modeli (ELECTRA-TR)

ESA	: Evrişimli Sinir Ağları (CNN)
EYK	: En Yakın Komşu (kNN)
GA	: Gradyan Artırmalı Öğrenme (GB / GBoost)
GDES	: Gradyan Ayırıştırılmış Gömme Paylaşımı (GDES)
GK	: Gradyan Kuvvetlendirme (GB)
GPT	: Üretken Ön-Eğitilmiş Transformer (GPT)
HESA	: Hibrit Evrişimsel Sinir Ağları (BiLSTM-CNN / CNN-LSTM)
HP	: Hiperparametre (HP)
HPO	: Hiperparametre Optimizasyonu (HPO)
HYB	: Hibrit Yöntemler (HYB)
IA-UKSB-TSA	: İnce Ayarlı UKSB-TSA (FT-LSTM-RNN)
KA	: Karar Ağacı (DT)
KDM	: Küçük Dil Modeli (SLM)
KG	: Kelime Gömme (WE)
K-Kare	: Ki-Kare Özellik Seçimi (CHI)
K-NN	: k-En Yakın Komşu (kNN / k-NN)
KT	: Kelime Torbası (BoW)
KTB	: Kapılı Tekrarlayan Birim (GRU)
KTY	: Kural Tabanlı Yöntemler (RB)
LORA	: Düşük-Dereceli Uyumlama (LORA)
LR	: Lojistik Regresyon (LR/LOGR)
MDM	: Maskelenmiş Dil Modellemesi (MLM)
MNB	: Çoğul Naif Bayes (MNB)
MÖ	: Makine Öğrenimi (ML)
NB	: Naif Bayes (NB)
NER	: Adlandırılmış Varlık Tanıma (NER)
ÖM	: Özellik Mühendisliği (FE)
PEFT	: Parametre-Verimli İnce Ayar (PEFT)
POS	: Sözcük türü etiketleme (POS)
PVIA	: Parametre-Verimli İnce Ayar (PEFT)
RO	: Rastgele Orman (RF)
RoBERTa	: Sağlamlaştırılmış ve Optimize Edilmiş BERT (RoBERTa)
STY	: Sözlük Tabanlı Yöntemler (LB)
TDA	: Tüketici Duyarlılık Analizi (CSA)
TF	: Terim Frekansı (TF)

TF-TDF	: Terim Frekansı–Ters Doküman Frekansı (TF-IDF)
TİA	: Tam İnce Ayar (FFT)
TÖ	: Topluluk Öğrenmesi (EL)
TR-BTweet	: Türkçe BERTweet (TR-BERTweet)
TRF	: Transformer (TRF)
TSA	: Tekrarlayan Sinir Ağı (RNN)
TSK-DVM	: Tek Sınıfa Karşı Destek Vektör Makineleri (OVA-SVM)
TTİA	: Talimat-Temelli İnce Ayar (IT)
TY	: Topluluk Yöntemleri (ENS)
UKSB	: Uzun Kısa Süreli Bellek (LSTM)
UKSB-ESA	: CNN–LSTM Hibrit (CNN-LSTM)
UKSB-TSA	: UKSB–TSA Hibrit (LSTM-RNN)
ÜYZ	: Üretken Yapay Zekâ (GAI)
XLM	: Çok Dilli Dil Modeli (XLM)
XLM-R	: XLM-RoBERTa (XLM-R)
XLM-T	: XLM-Twitter (XLM-T)
XLNet	: XLNet Oto-regresif Model (XLNet)
YDÖ	: Yarı-Denetimli Öğrenme (SSL)
YSA	: Yapay Sinir Ağı (ANN)



ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması	19
Çizelge 3.1 : Popüler Ollama büyük dil modelleri	40
Çizelge 3.2 : Karmaşıklık matrisi	51
Çizelge 4.1 : Veri setleri analizi	59
Çizelge 4.2 : Ürün yorumları	62
Çizelge 4.3 : Ürün soruları ve cevapları	63
Çizelge 4.4 : Yorum, puan ve etiket dağılımları.....	64
Çizelge 4.5 : Yorum etiket düzenlemeleri	66
Çizelge 4.6 : Nihai yorumlar	74
Çizelge 4.7 : Eğitim, doğrulama ve test veri seti dağılımları	79
Çizelge 4.8 : Eğitim, doğrulama ve test veri seti ile duygu sınıfı dağılımları	79
Çizelge 4.9 : Optuna optimizasyonu denemelerinde kullanılan alt veri seti	80
Çizelge 4.10 : Temel modeller	82
Çizelge 4.11 : Optuna HPO denemeleri epoch düzeyi	86
Çizelge 4.12 : Optuna HPO denemeleri	87
Çizelge 4.13 : Doğrulama veri seti ile önerilen modellerin performans analizi	92
Çizelge 4.14 : Test veri seti ile önerilen modellerin performans analizi	93
Çizelge 4.15 : Doğrulama veri seti ile önerilen modellerin sınıf bazlı performans analizi.....	94
Çizelge 4.16 : Test veri seti ile önerilen modellerin sınıf bazlı performans analizi ..	95
Çizelge 4.17 : Doğrulama veri seti ile önerilen modelin performans analizi	97
Çizelge 4.18 : Önerilen modelin epoch adım raporları.....	97



ŞEKİL LİSTESİ

Sayfa

Şekil 3.1 : BERT ön eğitim ve ince ayar süreçleri (Devlin ve diğ., 2019).....	33
Şekil 3.2 : Parametre verimli ince ayar (George, 2024)	43
Şekil 3.3 : Klasik optimizasyon yöntemleri (Behera, 2024).....	47
Şekil 3.4 : Klasik ve bayes optimizasyon karşılaştırma (Passos ve diğ., 2022)	48
Şekil 4.1 : Yorumlardaki kelime sayıları dağılımı.....	60
Şekil 4.2 : Ürün detay	61
Şekil 4.3 : Tüm yorumların puan dağılımı	64
Şekil 4.4 : Etiketlenen yorumların duygu etiketi dağılımı.....	67
Şekil 4.5 : Etiketlenen yorumların kutu grafik analizi.....	67
Şekil 4.6 : Nihai veri seti duygu etiketi dağılımı.....	74
Şekil 4.7 : Pozitif yorumların kelime yoğunluk bulutu	76
Şekil 4.8 : Negatif yorumların kelime yoğunluk bulutu	77
Şekil 4.9 : Nötr yorumların kelime yoğunluk bulutu.....	78
Şekil 6.1 : Ürün öneri sistemi arayüz tasarımı çalışması.....	102



TÜRKÇE E-TİCARET YORUMLARINDA TRANSFORMER TABANLI DUYGU ANALİZİ

ÖZET

Bu çalışma, önceden eğitilmiş Transformer mimarilerine dayalı ince ayarlı sınıflandırma modelleri geliştirerek ve Optuna hiperparametre optimizasyonu uygulayarak Türkçe e-ticaret ürün yorumları üzerinde duygu analizi performansını artırmayı amaçlamaktadır. Bu kapsamda sırasıyla veri setinin hazırlanması, model eğitimleri, sonuçların değerlendirilmesi ve önerilen modellerin yayınlanması süreçleri yürütülmüştür.

Çalışma kapsamında hazır veri seti kullanılmamış olup; sistematik bir işlem hattı aracılığıyla veri madenciliği, etiketleme ve kapsamlı ön işleme adımları ile güncel ve alan özgü veri seti oluşturulmuştur. Veri kalitesini ve tutarlılığını artırmak amacıyla iş akışına UFAL normalizasyonu, türkçe karakter düzenlemeleri, emoji dönüştürme, url etiket ve email gibi gürültülerin temizlenmesi, harf tekrarlarının azaltılması, noktalama, işaret ve boşluk düzenlemeleri gibi süreçler yürütülmüştür. Ayrıca Türkçe dışı, tekrar eden ve alan dışı yorumlar da filtrelenmiştir. Sonrasında yorum duygu etiketleri Gemma ve LLaMA gibi yerel büyük dil modellerinden yararlanılarak gözden geçirilmiştir. Nihai veri seti 50.160 etiketli yorum ile neticelendirilmiş olup model eğitimleri sürecinde bağlam daha anlamlı ve kararlı hale getirilmiştir.

Hazırlık sürecinin tamamlanmasının ardından son hâline getirilen veri kümesi; BERT, ModernBERT, RoBERTa, TurkishBERTweet, ELECTRA ve DeBERTa gibi güncel Transformer tabanlı temel modellerin hem varsayılan hiperparametrelerle hem de makro-F1 skorunu en üst düzeye çıkarmaya yönelik Optuna tarafından optimize edilen hiperparametre değerleri ile tam ince ayarlı eğitimleri gerçekleştirilmiştir.

Deneysel bulgular, Optuna tabanlı hiperparametre optimizasyonu uygulanarak eğitilen modellerin, varsayılan parametrelerle eğitilen modellere kıyasla makro-F1 skorunda yaklaşık %3–%7 oranında iyileşme sağladığını göstermektedir. Test veri setinde en yüksek performans, Optuna ile optimize edilen tam ince ayarlı dbmdz/bert modeli tarafından elde edilmiş olup, model %64,17 makro-F1 ve %91,69 doğruluk değerlerine ulaşmıştır. Buna karşın doğrulama veri setinde en başarılı sonuç, Optuna ile optimize edilen tam ince ayarlı Trendyol/tyroberta modeli tarafından elde edilerek %65,48 makro-F1 ve %92,51 doğruluk skorlarıyla kaydedilmiştir. Transformer bazlı modellerin klasik sınıflandırma algoritmalarına kıyasla daha yüksek Macro F1 sağladığı gözlemlenmiştir. Doğrulama ve test sonuçları arasındaki farkların düşük olması, modellerin genellenebilirliğinin yüksek olduğunu göstermektedir.

Genel olarak bu çalışma, modern Transformer mimarilerinin Türkçe e-ticaret bağlamında duygu analizi görevlerinde etkinliğini deneysel olarak ortaya koymakta ve hiperparametre optimizasyonunun model performansı üzerindeki belirleyici etkisini vurgulamaktadır. Yeniden üretilebilirliği sağlamak ve araştırma topluluğuna açık bir kaynak sunmak amacıyla, geliştirilen modeller Hugging Face platformunda kamuya açık şekilde yayımlanmıştır.



TRANSFORMER-BASED SENTIMENT ANALYSIS ON TURKISH E-COMMERCE REVIEWS

SUMMARY

With the rapid expansion of digitalization, e-commerce platforms have become one of the most influential sources of information shaping consumers' purchasing behaviors. User-generated product reviews not only serve as a reference point for potential buyers but also provide critical insights for manufacturers and sellers in terms of measuring customer satisfaction, managing product development processes, and gaining a competitive advantage. In this context, the automatic processing of large-scale and unstructured textual data to extract meaningful insights has brought sentiment analysis—widely studied within the field of Natural Language Processing (NLP)—to the forefront. However, e-commerce reviews inherently exhibit a noisy structure due to spelling errors, grammatical inconsistencies, slang expressions, and the extensive use of emojis, which significantly limits the effectiveness of traditional classification algorithms.

Turkish, with its agglutinative morphological structure and rich word-formation capacity, presents unique challenges for sentiment analysis compared to widely studied languages such as English. The ability of word stems to remain unchanged while acquiring numerous suffixes substantially increases vocabulary size and leads to data sparsity issues. Although the literature indicates a growing number of studies on Turkish sentiment analysis, a clear research gap remains—particularly regarding the systematic and experimental investigation of handling class imbalance in e-commerce-specific datasets and the impact of hyperparameter optimization on model performance. Motivated by this gap, this thesis aims to enhance the effectiveness of modern Transformer architectures on Turkish e-commerce data and to empirically demonstrate the role of hyperparameter optimization (HPO) in this process.

One of the most critical factors determining the success of sentiment analysis is the quality of the data. Accordingly, instead of relying on an existing dataset, the dataset used in this study was constructed from scratch to accurately reflect real-world scenarios. The data mining process was conducted on the “Sunglasses” category of a JavaScript-based e-commerce platform using Python-based Selenium and BeautifulSoup libraries. As a result of crawling 11,014 products across 242 brands, a total of 159,552 raw user reviews were collected. The raw data was found to be highly imbalanced (predominantly positive) and noisy due to the nature of e-commerce reviews, which often contain spelling errors, emojis, and repeated characters. To prepare the dataset for model training, systematic labeling, preprocessing, and filtering strategies were applied.

To mitigate class imbalance during the labeling process, the majority of neutral and negative reviews and a subset of positive reviews were selected, resulting in a labeled dataset of 60,297 samples. A comprehensive preprocessing pipeline was implemented to improve data quality and ensure suitability for model training. In particular, the

UFAL (MultiLexNorm) model—winner of the W-NUT 2021 shared task and based on ByT5 character-level normalization—was employed to correct frequent spelling errors commonly observed in e-commerce product reviews. Due to limitations in the pretraining data of the UFAL model, Turkish character corrections were further supported using the Zemberek Deasciify library. Emojis, which play a crucial role in sentiment expression, were not removed but instead converted into textual representations to preserve their semantic meaning. Additional preprocessing steps included the removal of URLs, tags, and email addresses; reduction of repeated characters; and normalization of punctuation, symbols, and whitespace. Furthermore, non-Turkish, duplicate, and domain-irrelevant reviews were filtered to further refine the dataset.

In addition to manual labeling, predictions were generated using local large language models (LLMs), namely Gemma3:4b and LLaMA3.2:3b, on the fully preprocessed dataset to minimize human error and enhance consistency. Manual labels, LLM predictions, and user ratings were jointly evaluated, resulting in the identification of 1,101 inconsistently labeled reviews. These samples were manually reviewed, and final sentiment labels (Positive, Negative, Neutral) were assigned. Following labeling, preprocessing, and filtering, the final dataset consisted of 50,160 labeled reviews, yielding a more stable and semantically coherent context for model training. The dataset was split into training (68%), validation (12%), and test (20%) sets, ensuring that class distributions were preserved across all subsets through stratified splitting.

After completing the dataset preparation phase, six Transformer-based base models—BERT, ModernBERT, RoBERTa, TurkishBERTweet, ELECTRA, and DeBERTa—were selected based on criteria such as architectural design, training strategies, case sensitivity (cased/uncased), Turkish language support, vocabulary size, tokenizer characteristics, model recency, popularity, and compatibility with social media and e-commerce language. Additionally, classical classification algorithms, including Complement Naive Bayes and k-Nearest Neighbors, were incorporated to provide a comprehensive performance comparison framework.

One of the key methodological strengths of this study lies in performing fine-tuning not only with default hyperparameter values but also through a systematic hyperparameter optimization process. In Transformer-based models, hyperparameters such as learning rate, batch size, warmup ratio, and weight decay have a substantial impact on performance. Unlike grid search or random search methods, Bayesian Optimization offers more efficient exploration by learning from previous evaluations, reducing computational cost, and incorporating early stopping strategies. Therefore, the Bayesian optimization-based Optuna framework was employed. For each model, 100 trials were conducted to identify the optimal hyperparameter combinations that maximize the Macro-F1 score, which was selected as the primary objective metric due to class imbalance. Final training was completed using these optimized hyperparameter values. All experiments were conducted locally on an NVIDIA GeForce RTX 3070 Ti GPU, and results were evaluated using both overall and class-level metrics.

Experimental results demonstrate that models trained with Optuna-based hyperparameter optimization achieve approximately 3–7% improvements in Macro-F1 scores compared to models trained with default hyperparameter values. The highest performance on the test dataset was achieved by the Optuna-optimized, fully fine-tuned Dbmdz/bert-base-turkish-128k-uncased model, attaining a Macro-F1 score of

64.17% and an accuracy of 91.69%. Conversely, the best performance on the validation dataset was obtained by the Optuna-optimized, fully fine-tuned Trendyol/tyroberta model, with a Macro-F1 score of 65.48% and an accuracy of 92.51%. The success of these two models (BERT- and RoBERTa-based) clearly highlights the advantages of large vocabularies in morphologically rich languages such as Turkish (BERT-128k) and domain-specific pretraining on e-commerce data (TyRoBERTa). Social media-oriented TurkishBERTweet and multilingual mDeBERTa models also achieved competitive results, albeit slightly below the leading models. Overall, Transformer-based models consistently outperformed classical classification algorithms in terms of Macro-F1 performance. The minimal discrepancy between validation and test results further indicates strong model generalizability.

Class-level analyses reveal that the models performed exceptionally well in identifying Positive (F1 > 95%) and Negative (F1 > 80%) classes. Notably, the high accuracy in detecting negative reviews is particularly valuable for organizations aiming to capture customer complaints effectively. However, identifying the Neutral class remained a challenging task across all models (F1 < 20%). This difficulty can primarily be attributed to class imbalance and the inherently high semantic ambiguity of neutral expressions.

In summary, this study empirically demonstrates the effectiveness of modern Transformer architectures in Turkish e-commerce sentiment analysis tasks and underscores the decisive role of hyperparameter optimization in improving model performance. The proposed models offer direct practical value for decision-making processes such as product management, customer experience analysis, and product assortment planning. Automating the consistent analysis of tens of thousands of customer reviews—previously requiring manual inspection—provides organizations with significant time and cost advantages while enabling faster and more accurate identification of customer expectations. Moreover, jointly analyzing sentiment distributions alongside product attributes, price levels, and user profiles facilitates data-driven decision-making in marketing strategies, product improvement initiatives, and portfolio optimization. Consequently, sentiment analysis transcends its traditional role as a customer satisfaction measurement tool and emerges as a strategic decision-support mechanism for product development, churn prediction, marketing strategy formulation, and competitive advantage creation.

To ensure reproducibility and provide an open resource for the research community, the proposed models have been publicly released on the Hugging Face platform. This initiative enables both the Turkish NLP community and industry practitioners to directly utilize or further develop the pretrained models in their own projects. Future work includes comparing different hyperparameter optimization techniques during model fine-tuning and addressing class imbalance by augmenting the dataset through synthetic data generation and advanced data mining approaches. Additionally, leveraging the developed models to perform sentiment inference on unlabeled datasets opens avenues for building recommendation systems based on customer sentiment.



1. GİRİŞ

Duygu analizi (DA), doğal dil işlemenin (DDİ) temel araştırma alanlarından biri olup, bireylerin metin üzerinden aktardıkları duygu, görüş veya değerlendirmeleri otomatik biçimde tespit etmeye yönelik bir sınıflandırma problemidir (Liu, 2012). Çevrimiçi incelemelerin hızla yayılmasıyla birlikte, büyük hacimli yapılandırılmamış metinsel veriyi karar vericiler için "eyleme geçirilebilir bilgiye" dönüştürülmesi yeteneği nedeniyle DA hayati bir araç haline gelmiştir.

Duygu analizi yöntemleri tarihsel olarak beş ana aşamada gelişmiştir. İlk dönemde ortaya çıkan kural tabanlı yöntemler, duygu sözlükleri ve dilsel göstergeler üzerinden puanlama yaparak temel bir yaklaşım sunmuş ancak bağlam, ironi ve çok anlamlılık gibi dilin karmaşık yönlerini yakalamada yetersiz kalmıştır. Bunu izleyen makine öğrenimi dönemi, metinlerin TF-IDF gibi sayısal temsillere dönüştürülmesiyle SVM ve NB gibi sınıflandırıcıların kullanılmasını sağlamış olsa da kelimeler arası bağlamsal ilişkileri modelleyememiştir. Derin öğrenmenin yükselişi ile birlikte Word2Vec, GloVe ve FastText gibi gömme teknikleri, ardından LSTM, GRU ve CNN gibi mimariler duygu analizinde önemli ilerlemeler sağlamıştır. Asıl dönüşüm ise Transformer mimarisinin ve bu mirari ile BERT, RoBERTa, ELECTRA gibi dil modellerinin alana girmesiyle yaşanmış; bu modeller bağlamı iki yönlü ve derin şekilde temsil ederek performans standartlarını yeniden tanımlamıştır. Günümüzde ise ChatGPT ve LLaMA gibi büyük dil modelleri, örneksiz öğrenme ve az örnekli öğrenme yetenekleri sayesinde duygu analizini minimal veriyle gerçekleştirilebilen, daha esnek ve genellenebilir bir yapıya kavuşturmuştur.

Bu tez çalışmasında, literatürde öne çıkan güncel Transformer tabanlı duygu analizi modelleri temel alınarak Optuna kullanımıyla hiperparametre optimizasyonu gerçekleştirilmiş ve bu doğrultuda ince ayarlı modellerin oluşturulması amaçlanmıştır. Veri hazırlama aşaması, veri madenciliği, etiketleme ve ön işleme adımlarının sistematik biçimde yürütülmesiyle tamamlanmış; sonrasında modeller hem varsayılan hiperparametre değerleriyle hem de Optuna tarafından belirlenen optimize edilmiş parametre değerleriyle eğitilmiştir. Elde edilen sonuçlar, genel ve sınıf bazlı

performans metrikleri üzerinden karşılaştırmalı olarak değerlendirilmiş; en başarılı performansı sergileyen modeller Hugging Face platformunda yayımlanarak araştırmacıların kullanımına açılmıştır.

Bu çalışmanın temel motivasyonu, yapay zekâ alanına duyduğum ilgi doğrultusunda, daha önce doğrudan deneyimim olmayan duygu analizi ve model geliştirme süreçlerinde kendime yeni bir öğrenme alanı açma isteğimden kaynaklanmaktadır. Agentic AI gibi güncel yaklaşımların hızla geliştiği bir dönemde, bu alanın önemli bir yapı taşı olan duygu analizinde uygulamalı deneyim kazanmak ve teknik yetkinliğimi artırmak benim için değerli bir fırsat sunmuştur. Diğer yandan, Türkçe ve sektör odaklı veri setlerinin sınırlı olması ile alan özelinde geliştirilen modern modellerin eksikliği, bu çalışmayı yalnızca kişisel bir öğrenme süreci olmaktan çıkarıp literatürdeki mevcut boşluğa katkı sunma yönünde bir motivasyona dönüştürmüştür. Bu nedenle veri madenciliği, veri ön işleme, etiketleme gibi temel aşamalardan başlayarak Transformer tabanlı modeller üzerinde hiperparametre optimizasyonu gerçekleştirmek hem güncel yöntemleri uygulamada deneyimlememe hem de alana özgü duygu analizi modellerinin geliştirilmesine katkı sağlamama imkân tanımıştır.

2. LİTERATÜR ARAŞTIRMASI

Literatürde yer alan çalışmalar üç ana başlık altında kapsamlı şekilde ele alınmıştır: Sistematik Literatür Taramaları, Ampirik Çalışmalar ve Metodoloji.

İlk olarak, duygu analizi kapsamında temel kavramları, kuramsal çerçeveleri ve güncel eğilimleri sistematik ve yapılandırılmış bir bakış açısıyla değerlendiren Sistematik Literatür Taramaları çalışmaları sunulmaktadır.

İkinci başlıkta ise gerçek dünyadan elde edilen veri setleri, deneysel yaklaşımlar ve uygulamalı yöntemlerle üretilmiş bulguları içeren Ampirik Çalışmalar incelenerek, teorik çerçevelerin pratik karşılıkları ortaya konulmaktadır.

Son olarak, tez kapsamındaki uygulama sürecinde kullanılan yöntemlerin bilimsel dayanaklarını açıklamayı amaçlayan Metodolojiye Yönelik literatür incelemeleri yer almaktadır.

Bu bölüm içersinde Sistematik Literatür Taramaları ve Ampirik Çalışmalar'a yer verilmekte olup, tez uygulaması yöntemlerinin bilimsel arka planını vermek üzere yapılan literatür incelemeleri ise Metodoloji bölümü altında yer almaktadır.

Bu bölümde incelenen literatür çalışmaları detaylıca sınıflandırılması gerçekleştirilmiş olup Çizelge 2.1'de gösterilmektedir.

2.1 Sistematik Literatür Taramaları

Duygu analizi alanındaki güncel eğilimleri, kullanılan yöntemleri ve araştırma boşluklarını bütüncül biçimde ortaya koyan sistematik literatür tarama çalışmaları incelenmektedir.

Jain ve diğ. (2021), konaklama ve turizm alanı odaklı kullanıcı yorumlarını kullanarak Tüketici Duyarlılık Analizi (TDA) için uygulanan Makine Öğrenimi (MÖ) tekniklerinin faydasını, kapsamını ve uygulanabilirliğini analiz etmeye odaklanmıştır. Çalışma, Ocak 2017'den Temmuz 2020'ye kadar yayınlanmış 68 makaleyi içermektedir. MÖ tekniklerinin TDA'nın üç temel alt alanında uygulandığı tespit

edilmiştir: Duygu Analizi Sınıflandırması (39 makale), Tahmine Dayalı Tüketici Tavsiye Kararları (15 makale), Sahte İnceleme Tespiti (14 makale). Duygu analizi kapsamında en sık kullanılan MÖ tekniklerinin Regresyon (16 makale), Destek Vektör Makineleri (SVM) (14 makale), Naif Bayes (NB) (11 makale) ve Lojistik Regresyon (LR) (11 makale) olduğu şeklinde raporlanmıştır. Nihayetinde araştırma, MÖ'nin TDA kapsamında entegrasyonunun ve hizmet sağlayıcı kuruluşların hizmet kalitesini iyileştirmede ve daha iyi tüketici politikaları oluşturmada bu entegrasyonu kullanmalarının faydalı olduğu sonucuna varmaktadır.

Tan ve diğ. (2023), doğal dil işlemenin metinleri pozitif, negatif ve nötr olmak üzere üç temel duygu kategorisine ayırmaya odaklanan duygu analizi alanındaki en son gelişmelere kapsamlı bir genel bakış sunmaktadır. Makale, duygu analizi sürecini (ön işleme, özellik çıkarımı ve sınıflandırma) detaylandırarak, çalışmalarını geleneksel makine öğrenimi (NB, DVM gibi), derin öğrenme (LSTM, CNN, BiLSTM gibi) ve daha iyi performans elde etmek için birden çok modeli birleştiren topluluk öğrenimi olmak üzere üç ana yaklaşıma ayırır. İncelenen yaygın veri setleri arasında Internet Movie Database (IMDb), Twitter US Airline Sentiment, Sentiment140 ve SemEval-2017 Task 4 bulunmaktadır. Bu veri setlerinin karmaşık dil yapısı, sınıf dengesizliği ve metinlerin kısalığı gibi zorlukları vurgulanmıştır. Son olarak, makale mevcut yöntemlerin kötü yapılandırılmış veya sarkastik metinleri ve 3 sınıflı duygu sınıflandırmasını ele almada zorlandığını belirtmekte; gelecekteki araştırmaların ise duygusal yoğunluğu yakalayan 5 sınıflı duygu analizi, duygu nicelendirmesi ve çapraz dilli modellerin geliştirilmesi gibi konulara odaklanması gerektiğini öne sürmektedir.

Ali ve diğ. (2025), duygu analizi alanındaki güncel araştırma çalışmalarını, yöntemleri, zorlukları ve gelecekteki eğilimleri kapsayacak şekilde sistematik literatür taraması yapmıştır. İncelenen temel teknikler; önceden tanımlanmış sözlükleri kullanan Sözlük Tabanlı Yaklaşımlar, etiketlenmiş eğitim verileri üzerinde çalışan Makine Öğrenimi (MÖ) algoritmaları ve otomatik olarak karmaşık örüntüleri öğrenen Derin Öğrenme (DÖ) modelleri ile bu yöntemlerin birleşimi olan Hibrit yaklaşımlardır (HYB). Çalışma, DÖ modellerinin, özellikle duygu polaritesi tespiti ve yön düzeyinde analiz gibi görevlerde, sözlük tabanlı ve geleneksel MÖ yöntemlerine kıyasla tutarlı bir şekilde daha iyi performans gösterdiğini ortaya koymuştur. Çalışma DA genel uygulama süreç akışı DA yaklaşım ve tekniklerin detaylı bir şekilde sınıflandırılması gibi önemli bulguları ortaya koymuştur. Sarkazmı tespit etmek ve farklı dillerle başa

çıkmak gibi hala çözülmeyi bekleyen önemli zorluklar bulunmakta olduğu ifade edilmektedir. Gelecekteki araştırmaların ise, DÖ modellerinin yorumlanabilirliğini artırmaya, metinle birlikte görüntü ve ses gibi çok modlu verileri entegre etmeye ve modellerdeki önyargıları önleme gibi etik sorunlara odaklanması önerilmektedir.

2.2 Ampirik Çalışmalar

Bu başlık altında, duygu analizi kapsamında uygulama çalışmalarına yer verilmektedir. Hangi veri seti ile hangi yöntem kullanılarak ne tarz sonuçlar elde edildiği incelenmektedir.

Lu ve diğ. (2010), 1 ile 5 yıldız arasında bir duygu skoru atanarak kullanıcı yorumlarının duygu şiddetini tahmin etme problemini çözmeyi hedeflemiştir. Çözüm için kullanılan yöntem, bir doğal dil işleme tabanlı melez yaklaşım üzerine kurulmuştur. İncelemelerdeki duygu şiddeti esas olarak sıfatlar ve zarflar aracılığıyla ölçülmüş, sık kullanılan zarfların güçleri sezgilere dayalı olarak manuel olarak -1 ile +1 arasında değerlerle etiketlenmiştir. Sıfatların duygusal gücü ise, arama motoru verileri ve ilerleyici kuralları kullanarak oluşturulan bir ilerleyici ilişki grafiği üzerindeki bağlantı analizi metoduyla hesaplanmıştır. Bu bilgi tabanlı yaklaşım, derecelendirmesi 1 ile 5 yıldız arasında olan Çince otel yorumlarından oluşan bir veri seti üzerinde uygulanmıştır. Elde edilen sonuçlar, önerilen yaklaşımın ikili duygu sınıflandırmasında etkili olduğunu ve sıfat gücünü dikkate almanın isabet oranını, bu gücü dikkate almayan yaklaşıma göre %5.7 oranında artırdığını (%71.65 Kesinlik) göstermiştir. Ayrıca, yorumlara 1 ile 5 yıldız arasında bir derece atanan beşli ölçeklendirme değerlendirilmesinde %38.88 Kesinlik ve %78.61 Uzaklık Keskinliği elde edilmiştir. Makalenin literatüre sağladığı temel katkı, duygu şiddetini hesaplamak için arama motoru verilerini ve ilerleyici kuralları entegre eden yeni bir bilgi tabanlı yaklaşım sunarak, çok ölçekli (5 yıldızlı) duygu derecelendirmesi yapılmasıdır.

Khan ve Baharudin (2011), metinleri nesnel ve öznel olarak ayırdıktan sonra, öznel cümleleri pozitif, negatif ve nötr olmak üzere üç etiketli sınıflandırmaya tabi tutmaktadır. Kullanılan veri seti, blog yorumları ve incelemelerden oluşmaktadır. Öznel cümlelerdeki görüş ifadelerinin anlamsal puanlarını belirlemek için SentiWordNet dizininden faydalanmaktadır. Cümlenin nihai duygusal ağırlığı hesaplanırken, tüm cümle yapısı, bağlamsal bilgiler, kelime anlamı ayrıştırması (word sense disambiguation) ve bağlamsal değer değiştiricileri (contextual valence shifter)

dikkate alınarak dinamik kalıp eşleştirme kuralları uygulanmaktadır. Deneysel sonuçlar, önerilen bu kural tabanlı ve alan bağımsız bir yaklaşımın o dönemki klasik makine öğrenmesi (MÖ) yöntemlerini geride bıraktığını ve doküman seviyesinde %87, cümle seviyesinde ise %83 doğruluk oranına ulaştığını göstermektedir. Bu makale, alan bağımsız doküman ve cümle seviyesi duygu sınıflandırması için kural tabanlı bir teknik sunarak ve üçlü polarite ayırımına odaklanarak döneminde literatüre önemli bir katkı sağlamıştır.

Jin ve diğ. (2012), e-ticaret siteleri (360buy ve Newegg) ve mikrobloglar (Sina ve Tencent mikroblogları) gibi çok kaynaklı ortamlardan toplanan Çince ürün yorumlarına duygu analizi uygulamayı ve otomatik olarak bilgi çıkarmayı hedeflemiştir. Makale, biri ürünün çeşitli yönlerine (aspect-base) dayalı ikili duygu sınıflandırması (pozitif/negatif) ve diğeri mikrobloglarda paylaşılan ürün yorumlarının üç sınıflı sınıflandırılması (pozitif, negatif, nötr) olmak üzere iki temel problemi ele almıştır. İkili sınıflandırma görevi için, 821 cep telefonu ürününe ait 663.537 pozitif ve 314.529 negatif yorumdan oluşan büyük bir veri seti kullanılmıştır. Bu görevde unigram özellikler, BOOL, TF ve TF-IDF gibi özellik ağırlıklandırma stratejileri ve Chi-square (CHI-5, CHI-20, CHI-100) ile boyut indirgeme teknikleri uygulanarak NB, KNN ve SVM gibi klasik makine öğrenmesi algoritmaları karşılaştırılmış; deneyler, BOOL ağırlıklandırma ile SVM sınıflandırıcısının en iyi performansı sağladığını göstermiştir. Üç sınıflı sınıflandırma için ise, toplanan veri seti üzerinden 729 pozitif, 345 negatif ve 1.026 nötr olarak manuel etiketlenmiş toplam 2.100 mesajlık bir mikroblog veri seti kullanılmıştır. Bu görevde Lineer Regresyon, İki Aşamalı SVM, Minimum Kesit Modeli optimizasyonu ve OVA-SVMs gibi yaklaşımlar denenmiş, bu yöntemler arasında OVA-SVMs %64,5 Doğruluk oranıyla en iyi sonucu vermiştir. Minimum Kesit optimizasyonu uygulanan İki Aşamalı SVM ise, $\alpha=0.2$ ve $t=0.5$ hiperparametreleri ile %64,0 doğruluk elde ederek OVA-SVMs'e yakın bir etki göstermiştir. Çalışmanın literatüre sağladığı temel katkı, önerilen duygu sınıflandırma çerçevesinde sadece açık (explicit) yönleri değil, aynı zamanda örtük (implicit) yönleri de keşfetme mekanizmasını dikkate almasıdır. Yazarlar, Çince ürün incelemelerinde bu tür bir örtük yön keşfi çalışmasının daha önce raporlanmadığını belirtmektedir. Çalışmanın bir kısmı halihazırda bir Çince Ürün İnceleme Madenciliği Sisteminde uygulamaya konulduğu ifade edilmektedir.

Vural ve diğ. (2013), eklemeli bir dil olan Türkçe metin belgelerinde denetimsiz duygu analizi için bir çerçeve sunmayı amaçlamış, temel problem olarak da kullanıcı tarafından üretilen metinlerin kutupluluğunu (pozitif veya negatif) tahmin etmeyi hedeflemiştir. Araştırmacılar, çerçevelerini popüler bir Türk film sitesi olan Beyazperde'den elde edilen, her biri 30.000 pozitif ve 30.000 negatif olmak üzere toplam 60.000 yorumdan oluşan geniş bir Türkçe film yorumları korpusu üzerinde ikili (pozitif-negatif) sınıflandırma uygulamasına odaklanarak değerlendirmişlerdir. Çalışmada kullanılan temel yaklaşım, sözlük tabanlı bir kütüphane olan SentiStrength'in Türkçeye özelleştirilmesi ve denetimsiz bir çerçeve içinde kullanılmasıdır. Önerilen çerçeve, girdiyi işleyerek polarite tahmini yapan bir ardışık düzenden (pipeline) oluşmaktadır; bu düzende ASCII karakter dönüştürücü (Türkçe karakter ikamelerini ele almak için), Zemberek kullanılarak morfolojik analiz, olumsuzlama işleyici (negation handler) ve özelleştirilmiş SentiStrength gibi modüller yer almaktadır. Özellikle olumsuzlama işleme adımında, SentiStrength'in varsayılan pencere boyutu yerine 3'lük bir pencere boyutu kullanılarak Türkçeye özgü optimizasyon yapılmıştır. Polarite tahmini için denenen üç farklı puanlama yaklaşımı (sentence-binary, sentence-max/min, word-sum) arasında, kelime düzeyindeki duygu puanlarını toplayan word-sum tekniği %75,90 Doğruluk oranıyla en iyi sonucu vermiştir. Çalışmanın literatüre sağladığı temel katkı, Türkçede doğal ve gürültülü metinlere uygulanabilen, tamamen denetimsiz, sözlük tabanlı bir duygu analizi çerçevesi önermesi ve bu çerçevenin, denetimli yöntemlerin performansına (daha önceki bir çalışmada raporlanan %85 doğruluk) yaklaşan makul bir doğruluk elde etmesidir.

Vinodhini ve Chandrasekaran (2014), e-ticaret alanında, Amazon'dan toplanan İngilizce ürün incelemelerinin ikili sınıflandırmasını (pozitif/negatif) hedeflemiştir. Toplam 500 incelemeden (365 pozitif, 135 negatif) oluşan bu veri setinde, standart yöntemlere alternatif olarak, hibrit makine öğrenimi yaklaşımları önerilmiştir. Yöntemler, Temel Bileşen Analizi (PCA) ile özellik azaltma ve Sınıflandırıcı Topluluğu (Ensemble of Classifiers) tekniklerini birleştirmektedir. Çalışmada PCA ile Torbalama (Bagged SVM) ve PCA ile Bayesien Güçlendirme (Bayesian Boosting) olmak üzere iki hibrit model, bireysel sınıflandırıcılar (LR, SVM) ile karşılaştırılmıştır. Modeller, unigram (Model I) ve n-gram kombinasyonu (Model II) özellik setleri üzerinde test edilmiştir. En iyi performansı, Model II için %92,5

Doğruluk, %98 Tamlik (Completeness) ve %93,3 Etkinlik (Effectiveness) oranlarıyla PCA + Bayesian Boosting yöntemi elde etmiştir. Makalenin temel katkısı, özellik azaltma ve topluluk öğrenme algoritmalarını birleştiren hibrit modeller sunması ve bu modellerin kalitesini geleneksel metrikler yerine dört farklı performans metriği kullanarak ölçmesidir.

Xia ve diğ. (2015), torba-kelime (BOW) modelinin duygu analizinde karşılaşılan önemli bir eksiklik olan kutupluluk kayması (polarity shift) problemini ele almayı amaçlayan Çift Duygu Analizi adlı basit ancak verimli bir denetimli çerçeve sunmaktadır. DSA, her bir orijinal inceleme için duygu-tersine çevrilmiş bir inceleme yaratarak veri genişletme yapar ve bu çiftleri Çift Eğitim (Dual Training) ve Çift Tahmin (Dual Prediction) algoritmalarında kullanarak sınıflandırıcıyı eğitir ve tahminleri iki taraflı değerlendirir. Modelin etkinliği, Amazon'dan toplanmış İngilizce ve Çince dillerinde, e-ticaret ürün incelemeleri üzerinde hem ikili (pozitif–negatif) hem de nötr yorumları içeren üç sınıflı sınıflandırma görevlerinde test edilmiştir. Veri setleri, ikili sınıflandırma için toplamda 8.000'er örnek içeren İngiliz ve Çin veri setlerinden; üç sınıflı sınıflandırma içinse yaklaşık 2.500 örnek içeren üç İngilizce veri setinden oluşmaktadır. Çalışmada NB, LR ve SVM gibi klasik makine öğrenimi algoritmaları kullanılmış; özellikle Çift Tahmin en iyi performansı sağlamıştır. DSA'nın en yüksek ortalama Doğruluğu SVM sınıflandırıcısı ile elde edilmiş, üç sınıflı sınıflandırma, temel yöntemleri geride bırakarak ortalama %73.6 Doğruluk oranı bildirmiştir. Makalenin temel katkısı, kutupluluk kayması sorununu çözmek için veri genişletme, çift eğitim ve çift tahmin mekanizmalarını kullanan bir çerçeve sunması ve harici sözlüğe bağımlılığı ortadan kaldırarak, dilden ve alandan bağımsız korpus tabanlı sözde-zıt anlamlılar sözlüğü (DSA-MI) geliştirerek modelin uygulanabilirliğini genişletmesidir.

Vinodhini ve Chandrasekaran (2016), e-ticaret uygulamalarında yaygın olan ve makine öğrenimi için zorluk teşkil eden dengesiz veri dağılımları altındaki duygu sınıflandırması problemine odaklanmıştır. Temel amaç, İngilizce dilinde yazılmış e-ticaret ürün incelemeleri veri setlerinde ikili sınıflandırma (pozitif/negatif) performansını artırmaktır. Çalışmada, Destek Vektör Makineleri (SVM) tabanlı bir topluluk algoritmasında değişiklik yapılarak Değiştirilmiş Torbalama (M-Bagging) yaklaşımı önerilmiştir. Bu hibrit makine öğrenimi yöntemi, aşırı ve az örnekleme yöntemlerini (SMOTE) entegre ederek dengesizlik (overfitting) sorununu gidermeyi

amaçlamaktadır. M-Bagging, standart SVM, Bagging, Boosting ve tekil örnekleme yöntemleri ile karşılaştırılmış ve farklı n-gram kombinasyonları üzerinde denenmiştir. Modellerin performansı, AUC (Area Under the ROC Curve) ve G-mean metrikleriyle değerlendirilmiştir. En iyi sonuçlar, en kapsamlı özellik setini kullanan Model III ile M-Bagging yöntemi tarafından elde edilmiştir: IBD1 veri setinde 0.89 AUC ve 0.90 G-mean. M-Bagging, yüksek dengesizlik oranı olan IBD2'de de diğer tüm yöntemlere göre daha iyi performans göstermiştir. Makalenin literatüre sağladığı temel katkı, örnekleme tekniklerini SVM temelli bir topluluk öğrenme (ensemble learning) çerçevesine entegre eden hibrit bir yaklaşım (M-Bagging) sunarak, özellikle yüksek dengesizlik oranına sahip veri setlerinde tahmin performansını önemli ölçüde iyileştirmesidir.

Ciftci ve Apaydin (2018), geleneksel makine öğrenimi (machine learning) yöntemlerinin (NB ve LR) torba-kelime (bag-of-words) modeli kullanması nedeniyle cümledeki kelime sırasını dikkate almama sınırlılığını aşmayı hedefleyerek, Türkçe duygu analizi (sentiment analysis) için derin öğrenme yaklaşımını önermiş ve bu yeni yaklaşımları klasik yöntemlerle karşılaştırmıştır. Çalışmanın temel amacı, doğal dil işleme (DDİ) alanındaki ve derin öğrenmedeki gelişmeleri, Türkçe alışveriş (hepsiburada) ve film (beyazperde) yorumları alanlarından toplanan yeni bir veri setine uygulamaktır. Elde edilen veri seti, 5 üzerinden puanlanmış yorumlardan oluşmakta olup, 2.5 puanın altındaki yorumlar negatif, üstündekiler pozitif kabul edilerek ikili sınıflandırma yapılmıştır. Veri seti, toplam 355.095 cümle içermekte olup, 282.670 pozitif ve 72.425 negatif cümle ile dengesiz bir dağılım sergilemektedir. Çalışmada, klasik yöntemler tf-idf vektörleştirme ile kullanılırken, derin öğrenme için kelime vektörleri ile beslenen ve kelime sırasını modelleyebilen RNN ve özel olarak LSTM birimleri kullanılmıştır. En iyi hiperparametreleri bulmak amacıyla Izgara Araması yöntemi uygulanmış, dengesiz veri sorununu gidermek için ise klasik yöntemlerde aşırı örnekleme, RNN'lerde ise maliyet fonksiyonunda farklı sınıf ağırlıkları kullanılmıştır. Karşılaştırmalı sonuçlar, LSTM modelinin en yüksek performansı sergilediğini ve test veri seti üzerinden %82,9 Doğruluğu, 0.86 Kesinlik ve 0.83 Duyarlılık değerlerini rapor ettiğini göstermiştir. Makalenin temel katkısı, Türkçe için derin öğrenme tabanlı duygu analizi uygulaması sunması ve 355 binden fazla yorum içeren yeni bir Türkçe online yorum veri seti oluşturmasıdır.

Santur (2019), 2 etiketli duygu analizi görevini ele almış ve bu sınıflandırma için Kapılı Yinelemeli Birim (GRU) mimarisinin uygunluğunu incelemeyi amaçlamıştır. Çalışmanın çözmeyi hedeflediği temel problem, GRU'nun, LSTM gibi diğer derin öğrenme algoritmalarına kıyasla daha az parametreye sahip olması ve CNN gibi yöntemlerde görülebilen kaybolan gradyan probleminden daha az etkilenmesi avantajını kullanarak, Türkçe duygu analizi görevindeki performansını ortaya koymaktır. Uygulama, Türkçe e-ticaret platformu Hepsiburada'dan alınmış kullanıcı yorumları veri seti üzerinde ikili sınıflandırma görevi ile gerçekleştirilmiştir. Veri seti, 243.497 satır yorum içermekte olup, bu yorumların %98'i pozitif (229.821) ve %2'si negatif (13.676) olmak üzere oldukça dengesiz bir dağılım sergilemektedir. GRU modeli, dört katmanlı (32, 16, 8, 4 birimli), Sigmoid aktivasyon fonksiyonlu ve Adam (.001) optimizasyon algoritması kullanılarak 16 epoch ve 256 yığın boyutuyla eğitilmiştir. Eğitilmiş GRU modeli ile doğrulama verileri üzerinde 0.955 Doğruluk, 0.971 Kesinlik, 0.980 Duyarlılık ve 0.976 F1 skoru değerleri rapor edilmiştir. Makalenin literatüre sağladığı temel katkı, Türkçe e-ticaret yorumları üzerinde GRU tabanlı bir duygu analizi uygulamasını sunması ve GRU'nun LSTM gibi daha karmaşık modellere kıyasla daha hızlı olması nedeniyle, hız ve benzer doğruluk sonuçları açısından Türkçe duygu analizi için uygun bir algoritma olduğunu göstermesidir.

Aydın ve Güngör (2021), Türkçe duygu analizi alanında denetimli yaklaşımlara aşırı odaklanılmasını ve dilin morfolojik yapısının göz ardı edilmesini ele almıştır. Temel amaç, zıt anlamlı kelime çiftlerine dayalı yeni denetimsiz (alandan bağımsız) ve yarı denetimli (alana özgü) yöntemleri geliştirerek, bu yaklaşımları denetimli modellerle birleştiren kapsamlı bir çerçeve sunmaktır. Çalışma, Türkçe film yorumları (20.244 yorum) ve Twitter veri seti (1.716 tweet) olmak üzere iki farklı metin stili üzerinde ikili sınıflandırma görevini ele almıştır. Kullanılan yöntemler, yeni denetimsiz/yarı denetimli yaklaşımlar, klasik makine öğrenimi sınıflandırıcıları (J-48, SVM, kNN, NB) ve derin öğrenme mimarileri (LSTM, CNN) ile morfolojik analiz ve özellik ağırlıklandırma tekniklerini içermektedir. Optimal katsayıları bulmak için denetimli ve denetimsiz yaklaşımların birleştirilmesi aşamasında ızgara araması kullanılmıştır. En iyi performansı, denetimsiz ve denetimli yaklaşımların birleştirilmesi (optimal katsayılar $cu=0.3$ ve $cs=0.7$) ve çoğunluk oylaması (majority voting) sonucu elde edilen yöntem sağlamıştır; bu yöntem film veri setinde %91,17 ve Twitter veri setinde %80,59 F1-skoru elde etmiştir. Çalışmanın temel katkısı, Türkçe için alandan

bağımsız duygu analizi yaklaşımını uygulayan ilk çalışma olması, denetimli, yarı denetimli ve denetimsiz yaklaşımları birleştiren bir çerçeve geliştirmesi ve Türkçe'nin eklemeli yapısını dikkate alarak morfolojik analizle elde edilen kısmi yüzey formlarının kullanımını önermesidir, bu sayede performans önemli ölçüde artırılmıştır.

Aktaş ve diğ. (2021), Türkçe duygu analizi alanında online yemek siparişi (yemeksepeti.com) yorumlarının sınıflandırılmasını hedeflemiş ve ilgili yorumlardaki duyguyu otomatik olarak algılayarak mümkün olan en yüksek doğruluk oranına ulaşmayı amaçlamıştır. Analizler, Türkçe dilinde, Yemeksepeti.com platformundan derlenmiş ve literatürde daha önce bulunmayan yaklaşık 676 bin yorumdan oluşan, dengeli bir veri seti üzerinde ikili sınıflandırma (pozitif/negatif) göreviyle gerçekleştirilmiştir (338 bin pozitif, 338 bin negatif). Çalışmada kullanılan yöntemler, klasik makine öğrenimi yaklaşımları (SVM, K-NN, NB, LR) ve derin öğrenme modelleri (Bidirectional CuDNNGRU, CuDNNGRU, CuDNNLSTM varyasyonları) olup, bu modeller kök bulma, kelime düzeltme ve çalışmanın yenilikçi katkısı olan klavye hatalarını düzeltmeye yönelik yöntem içeren üç farklı doğal dil işleme tekniğinin varyasyonlarıyla oluşturulmuş altı farklı veri seti üzerinde karşılaştırılmıştır. Hiperparametre optimizasyonu, Naif Bayes ve Lineer Regresyon için GridSearchCV fonksiyonu ile yapılırken, Yapay Sinir Ağı modeli için ise sistematik ve manuel olarak gerçekleştirilmiştir. En iyi sonuç, tam ön işleme tabii tutulmuş varsayılan veri seti üzerinde SVM algoritması tarafından %87,24 Doğruluk değeri ile elde edilmiştir. Makalenin temel katkıları; Türkçe için etiketli ve homojen yapıda, daha önce bulunmayan büyüklükte (676 bin yorum) yeni bir veri seti oluşturulması ve uygulanan üç farklı ön işleme yönteminin (Sözcük Kökü Bulma, Kelime Düzeltme ve Klavye Hata Düzeltme) toplu halde uygulandığında, yöntemin uygulanmadığı duruma göre doğrulukta yaklaşık %5'lik bir artış sağladığının gözlemlenmesidir.

Singh ve diğ. (2022), derin öğrenme ağlarının yüksek doğruluk seviyesine ulaşmak için nasıl adapte edileceği ve özellikle özellik ağırlıklandırmasının nasıl iyileştirileceği temel problemini çözmeyi amaçlamıştır. Bu amaçla, İngilizce Twitter verileri üzerinden COVID-19 incelemelerinin duygu analizi için bir derin öğrenme yaklaşımı sunulmuştur. Çalışmada kullanılan veri seti, toplam 179.108 COVID-19 ilişkili tweetten oluşmaktadır. Bu tweetler, pozitif (%45), nötr (%30) ve negatif (%25) şeklinde üçlü sınıflandırma ile etiketlenmiştir. Ayrıca, eğitim sırasında duygusal

bağlamı artırmak amacıyla üzgün, neşe, korku ve öfke olmak üzere dört sınıf etiketi de kullanılmıştır. Önerilen ana yöntem, kelime dizilerini öğrenme yeteneğine sahip olan LSTM-RNN üzerine kurulmuş ve dikkat katmanları kullanılarak özellik ağırlıklandırması ve semantik sıralama ile geliştirilmiştir. Karşılaştırma amaçlı olarak NB, RF, SVM ve LR gibi geleneksel makine öğrenimi yaklaşımları da kullanılmış ve bu klasik sınıflandırıcıların parametreleri ızgara araması kullanılarak ayarlanmıştır. En iyi performansı, dört dikkat katmanı ve Leaky ReLU aktivasyon fonksiyonu kullanılarak geliştirilen önerilen derin öğrenme yaklaşımı sergilemiştir. Bu yöntem, diğer mevcut yaklaşımlara kıyasla önemli bir iyileşme ile %84,56 Doğruluk, %82,34 Kesinlik, %82,12 Duyarlılık ve %81,23 F-skoru değerlerini rapor etmiştir. Makalenin literatüre sağladığı temel katkı, COVID-19 ile ilgili İngilizce tweetler üzerinde dikkat mekanizmasıyla geliştirilmiş LSTM-RNN tabanlı bir derin öğrenme yaklaşımı sunması ve bu sayede özellik ağırlıklarının semantik ilişkiye göre optimize edilerek performansın önemli ölçüde artırılmasıdır, bu da önceki yaklaşımlara kıyasla performansta belirgin bir iyileşme sağlamıştır.

Ben ve diğ. (2023), artan çevrimiçi satışlar bağlamında müşteri yorumlarından doğru duygu kutupluluklarını çıkarılmasının önemini vurgulamış ve geleneksel makine öğrenimi modelleri ile derin öğrenme ve hibrit modellerinin performansını karşılaştırmalı olarak analiz etmeyi amaçlamıştır. Çalışma, İngilizce dilinde, Amazon e-ticaret platformundan toplanmış, 36 milyon yorum içeren (18 milyonu pozitif) büyük bir veri seti üzerinde ikili sınıflandırma (pozitif/negatif) göreviyle gerçekleştirilmiştir. Yöntem olarak, özellik çıkarımı için TF-IDF vektörleştirilmesi kullanılmış; SVM, RF ve NB gibi klasik MÖ algoritmaları ile CNN, LSTM, Bi-LSTM gibi DÖ modelleri ve CNN-LSTM, CNN-BiLSTM gibi hibrit yaklaşımlar test edilmiştir. Yapılan deneylerde en iyi performansı, diğer tüm yöntemleri geride bırakarak CNN modeli sergilemiş ve %85 Doğruluk elde etmiştir. Makalenin temel katkısı, çevrimiçi ürün yorumları üzerinde yapılan bu karşılaştırmalı analizin, CNN başta olmak üzere derin öğrenme modellerinin, verinin bölgesel özelliklerini daha iyi yakalama kapasitesi sayesinde klasik MÖ algoritmalarına göre üstünlük gösterdiğini ortaya koymasındır.

Kumar ve diğ. (2023), yöntemlerin sınırlamalarını aşmayı ve müşteri duyarlılıklarını doğru bir şekilde kategorize etmek amacıyla sağlam ve otomatik bir duygu sınıflandırma modeli geliştirmeyi amaçlamıştır. Çalışma, Amazon ürün yorumlarından oluşan ve pozitif, negatif veya nötr duyarlılık etiketi içeren 50.000 örneklilik bir veri

seti üzerinde üçlü sınıflandırma göreviyle gerçekleştirilmiştir. Yöntemler olarak, araştırmacılar hem geleneksel makine öğrenimi sınıflandırma yöntemlerini LSVM, RF, MNB, BNB ve LR hem de derin öğrenme modellerini LSTM, Bi-LSTM ve BERT kullanmışlardır. Analiz öncesinde, veri gürültüsünü gidermek ve model doğruluğunu optimize etmek amacıyla küçük harfe dönüştürme, bağlantı kaldırma, durak kelime (stop word) kaldırma, kök bulma (stemming) ve lematizasyon gibi kapsamlı metin ön işleme adımları uygulanmıştır. Yapılan deneysel değerlendirmeler sonucunda, kullanılan modeller arasında BERT modeli, %92,25 Doğruluk puanıyla en yüksek genel performansı elde etmiştir. Bi-LSTM modeli ise %90,7 Doğruluk, %74 Kesinlik ve %76 Duyarlılık değerleriyle ikinci en yüksek performansı göstermiş ve LSTM modelini (%88 Doğruluk) geride bırakmıştır. Makalenin literatüre sağladığı temel katkı, Amazon yorumları bağlamında ileri düzey BERT ve Bi-LSTM modellerinin üstün performansını ampirik olarak kanıtlaması ve elde edilen bulgularla tüketici duyarlılıklarına dair değerli bilgiler sağlayarak karar alma süreçlerini ve çevrimiçi alışveriş deneyimini iyileştirmeye katkıda bulunmasıdır.

Alawi ve Bozkurt (2024), Türkçe gibi karmaşık morfolojik yapıya sahip bitişken dillerde duygu analizinin geleneksel yöntemlerle etkili bir şekilde ele alınamayan zorluklarını çözmeyi ve Türkçe kısa metin duygu analizi alanında derin öğrenme modelleri ile farklı gömme (embedding) tekniklerinin etkinliğini sistematik olarak araştırmayı amaçlamıştır. Çalışmanın deneysel kısmı, Türkçe dilinde, üniversite ve eğitim kalitesi hakkındaki Twitter (X) gönderilerinden oluşan dengesiz Türk Yükseköğretim Veri Seti (THED) (17.793 örnek) ve otel yorumlarından oluşan dengeli Otel Yorumları Veri Seti (HRD) (11.600 örnek) olmak üzere, her ikisi de ikili sınıflandırma (pozitif–negatif) etiketlerine sahip iki veri seti üzerinde yürütülmüştür. Kullanılan yöntemler, Word2Vec, GloVe, FastText gibi önceden eğitilmiş kelime gömme teknikleri ile iki karakter seviyesi gömme yöntemi CIE ve COE ve bu iki seviyeyi birleştiren hibrit yaklaşımları (CIE-WE, COE-WE) kapsamıştır. Bu gömme teknikleri, LSTM, CNN, Bi-LSTM ve CNN-BiLSTM, CNN-LSTM, BiLSTM-CNN gibi yenilikçi hibrit mimariler dahil altı farklı derin öğrenme modeli ile birlikte kullanılmış; modellerin en iyi performansı için hiperparametreler bayesian optimizasyon tekniği olarak Adam Optimizasyon kullanılarak ayarlanmıştır. Elde edilen sonuçlar, FastText gömme tekniğinin ve karakter ile kelime seviyelerini birleştiren hibrit yaklaşımların üstünlüğünü göstermiştir. Tut-bırak geçerlemede,

CNN-LSTM modeli FastText ile HRD veri setinde en yüksek F1 Skoru (0.9543) değerini rapor ederken, çapraz doğrulama ile CNN-LSTM modeli yine FastText ile 0.913 ± 0.013 F1 skoru elde etmiştir. Makalenin literatüre sağladığı temel katkılar; Türkçe metin temsili için daha önce keşfedilmemiş karakter seviyesinde tek-sıcak kodlama (COE) tekniğini önermesi ve karakter ile kelime seviyesini birleştiren çift yollu hibrit mimarinin, karmaşık morfolojik yapıya sahip dillerde duygu analizi performansını önemli ölçüde artırdığını göstermesidir.

Latha ve Rao (2024), e-ticaret platformlarında kullanıcı memnuniyetini artırmak amacıyla, geleneksel tavsiye sistemlerinin karşılaştığı soğuk başlangıç ve veri seyrekliği gibi sınırlamaları aşmayı ve büyük hacimli kullanıcı yorumları üzerindeki duygu analizinden yararlanarak otomatik bir ürün tavsiye sistemi geliştirmeyi temel problem olarak ele almıştır. Çalışma, İngilizce dilinde, Amazon Ürün Yorumları (APR) veri seti üzerindeki toplam 60.000 müşteri yorumu (25.627 pozitif, 19.567 negatif, 14.806 nötr) ile üçlü sınıflandırma görevi üzerinde gerçekleştirilmiştir. Problem, öncelikle durak kelime kaldırma, kök bulma ve lematizasyon gibi veri gürültüsü giderme teknikleri ve ardından TF-IDF tekniği kullanılarak metin verilerinin sayısal vektörlere dönüştürülmesi ile çözülmüştür. Bu vektörler daha sonra, ana yöntem olarak, özellik vektörlerinin boyutlarını azaltmak ve bağlamsal anlamayı geliştirmek amacıyla kelime gömme katmanına önceden eğitilmiş gömme yöntemlerini (skip-gram ve GloVe) ve kullanıcılar arasındaki benzerliğe dayalı tavsiyeyi kolaylaştırmak için Tekil Değer Ayrışımı (SVD) katmanını entegre eden Modifiye Evrişimsel Sinir Ağı (MCNN) adlı yeni bir derin öğrenme modeline beslenmiştir. MCNN modeli, karşılaştırmalı modellere kıyasla üstün performans elde etmiş ve APR veri setinde ortalama %97,40 Doğruluk ve %81,20 F1 skoru rapor etmiştir. Makalenin literatüre sağladığı temel katkı, birleştirilmiş kelime gömmelerin topluluk etkisinden ve SVD katmanının veri sadeleştirmesinden yararlanarak tavsiye kalitesini ve güvenilirliğini artıran, boyut azaltma sağlayan ve geleneksel yöntemlerin sınırlamalarını aşan yeni MCNN mimarisinin sunulmasıdır.

Özdemir ve diğ. (2024), duygu analizinde, geleneksel yöntemlerin sınırlamalarını aşmayı hedefleyerek Türkçe otel yorumları için Büyük Dil Modellerinin uygulamasını araştırmayı amaçlamıştır. Çalışmanın deneysel kısmı, orijinalde İngilizce olan bir korpusun Nöral Makine Çevirisi (NMT) kullanılarak Türkçe'ye otomatik olarak çevrilmesiyle oluşturulan 5.000 yorumdan oluşan bir veri seti üzerinde

gerçekleştirilmiştir. Bu veri seti, otel rezervasyon şirketlerinden toplanmış olup, kullanıcı puanları (0 ile 10 arası) 5'ten yüksekse pozitif (1), aksi takdirde negatif (0) olarak kabul edilerek ikili sınıflandırma etiketlemesi yapılmıştır. Çalışmada kullanılan temel yöntem, UL2 çerçevesine dayanan, son teknoloji ürünü Türkçe Dil Modeli olan TURNA'nın duygu sınıflandırma görevinde ince ayarlanması olmuştur. Bu modelin eğitiminde 'adamw' optimizasyon algoritması kullanılmış ve karşılaştırma için klasik BERT modeli de uygulanmıştır. Yapılan değerlendirmeler sonucunda, TURNA modelinin geleneksel makine öğrenimi algoritmaları ve diğer ileri düzey modellere kıyasla üstün performans sergilediği kanıtlanmıştır. TURNA modeli, %99,4 Doğruluk, %99,10 F1 skoru, %98,80 Kesinlik ve %99,4 Duyarlılık değerlerini rapor ederek en iyi performansı elde etmiştir. Makalenin literatüre sağladığı temel katkı, Türkçe otel yorumları için 5.000 örneklilik yeni bir veri seti oluşturup bunu herkese açık hale getirmesi ve TURNA gibi dile özgü LLM'lerin duygu sınıflandırma görevlerinde diğer gelişmiş modellere (klasik BERT'in yaklaşık %80 doğruluk oranı gibi) kıyasla önemli ölçüde daha yüksek doğruluk sağlayarak Türkçe duygu analizi metodolojilerinin gelişimine ve yerel dil modellerinin potansiyelini vurgulamasına katkıda bulunmasıdır.

Çakıcı ve diğ. (2025), Türkçe duygu analizi alanında karşılaşılan sınırlı veri seti bulunabilirliği ve farklı çalışmalar arasında aynı veri setlerinin yeniden kullanımından kaynaklanan çeşitli sonuçlar problemini ele almayı amaçlamıştır. Bu hedefle, 2012 ile 2025 yılları arasında yayımlanmış 78 ilgili çalışma sistematik olarak incelenmiş, bu çalışmalardan 80'den fazla veri seti çıkarılmış ve kapsamlı bir duygu analizi taksonomisi kullanılarak etiketlenmiştir. Çalışmanın deneysel kısmı, Türkçe dilinde, dört yaygın kullanılan veri seti (TMRD, Tmpr, Twt ve Humir) üzerinde, dört adet transformer tabanlı dil modeli XLM-T, BERTurk, TSAM ve TurkishBERTweet kullanılarak gerçekleştirilmiştir. Bu modeller hem üçlü hem de ikili sınıflandırma formatlarında test edilmiştir. Ampirik sonuçlar, model performansının veri seti özelliklerine (alan, denge ve dil yapısı) bağlı olarak önemli ölçüde değiştiğini göstermiştir. İkili sınıflandırma görevlerinde, XLM-T modeli, Twt veri setinde %92 Doğruluk ve %95 F1 skoru ile en yüksek performansı göstermiş, TSAM modeli ise Humir veri setinde neredeyse mükemmel %97 Doğruluk ve F1 skoru elde etmiştir. Çalışmanın literatüre sağladığı temel katkı, Türkçe duygu analizi araştırmalarına rehberlik etmek üzere sistematik olarak derlenmiş ve herkese açık merkezi bir veri seti

deposu (GitHub) oluřturması, yayımlanmış 78 alıřmanın taksonomik sınıflandırılmasını sunması ve modern transformer tabanlı modellerin, önceki geleneksel makine öğrenimi yöntemlerine göre önemli performans artışları sağladığını göstermesidir.

İncidelen ve Aydođan (2025), Türke gibi karmařık morfolojik yapıya sahip bitiřken dillerde duygu analizinin geleneksel yöntemlerle karřılařtıđı zorlukları ele almayı ve bu alanda güçlü modeller geliřtirmek amacıyla transformer tabanlı dil modellerinin performansını karřılařtırmalı olarak incelemeyi amalamıřtır. alıřmanın deneysel kısmı, Türke dilinde, e-ticaret platformlarındaki kullanıcı yorumlarından oluřan TRSAv1 veri seti üzerinde gerekleřtirilmiřtir. TRSAv1 veri seti, toplam 150.000 metinden oluřmakta olup, pozitif, negatif ve nötr olmak üzere üçlü sınıflandırma etiketlerine sahiptir ve her sınıf 50.000 yorum ierecek řekilde dengeli bir dađılıma sahiptir. alıřmada, altı adet transformer tabanlı modelin ince ayarlanması kullanılmıřtır: BERTurk 32k, BERTurk 128k, ELECTRA Turkish Small, ELECTRA Turkish Base (tek dilli) ile XLM-RoBERTa ve mBERT (ok dilli). İnce ayar sürecinde adil bir karřılařtırma sađlamak amacıyla hiperparametre optimizasyonu yapılmadan, Epoch 3, Batch Size 32 ve Learning Rate $3e^{-5}$ gibi özdeř hiperparametre deđerleri uygulanmıřtır. Yapılan deđerlendirmeler sonucunda, Türke iin özelleřtirilmiř tek dilli modellerin, ok dilli modellere kıyasla daha iyi performans sergilediđi görölmüřtür. En iyi performansı, %83,69 Doğruluk ve %83,65 F1 skoru ile BERTurk 32k modeli elde etmiřtir; bu model, özellikle nötr sınıfta diđer modellere göre daha iyi performans göstermiřtir. Makalenin temel katkısı, dengeli ve temsili bir Türke veri seti (TRSAv1) üzerinde Dil Modellerinin karřılařtırmalı performansını ortaya koyması, Türkenin dilsel yapısına daha duyarlı olan tek dilli modellerin ok dilli modellere karřı üstünlüğünü kanıtlayarak morfolojik aıdan zengin, düşük kaynaklı diller iin ince ayar stratejilerine yönelik öneriler sunmasıdır.

Teke ve diđ. (2025), e-ticaret platformlarındaki Türke ürün yorumlardaki duygu durumunu analiz etmeyi ve hatalı tahminleri en aza indirmeyi hedeflemiřtir. Deneysel alıřmalar, Trendyol e-ticaret sitesinden Selenium aracılıđıyla altı farklı kategoriden (Bilgisayar, Telefon, Ayakkabı, Giyim, Kozmetik, Spor ve Aık Hava) ekilen ok alanlı ürün incelemeleri üzerinde gerekleřtirilmiřtir. Orijinal veri seti, pozitif, negatif ve nötr etiketlerinden oluřan üçlü sınıflandırma yapısına sahip 73.392 yorumdan oluřmaktadır; ancak sınıf dađılımının dengesiz olması nedeniyle, küme tabanlı örnek

azaltma yöntemi uygulanarak 28.283 örneklilik dengelenmiş bir veri seti de oluşturulmuştur. Çalışmada kullanılan yöntemler, geleneksel makine öğrenimi modellerini SVM, RF, NB, LR, k-NN, GB ve DT ve transformer tabanlı dil modellerini BERTurk, mBERT, XLNet ve DistilBERT kapsamaktadır. Geleneksel modeller için özellik çıkarımı TF-IDF vektörleştirilmesi ile yapılmış, transformer modelleri için ise overfitting'i önlemek amacıyla eğitim beş epoch ile sınırlandırılmıştır. Yapılan kapsamlı deneyler sonucunda, BERTurk modeli, tüm geleneksel MÖ yöntemlerine ve diğer transformer tabanlı modellere kıyasla en başarılı model olarak belirlenmiştir. BERTurk, orijinal veri setinde %96 Doğruluk ve %96 F1 skoru ve alt örneklenmiş veri setinde ise %93 doğruluk ve %93 F1 skoru ile en yüksek performansı rapor etmiştir. Geleneksel modeller arasında ise SVM, her iki veri setinde de en yüksek performansı sergilemiştir (%89 orijinal, %84 alt örneklenmiş). Makalenin literatüre sağladığı temel katkı, çok alanlı, yeni bir Türkçe e-ticaret ürün incelemesi veri setini oluşturup yayımlaması ve BERTurk gibi modern transformer tabanlı modellerin, karmaşık metin yapısını daha iyi analiz etme yeteneği sayesinde, geleneksel makine öğrenimi yaklaşımlarına göre önemli ölçüde üstünlük sağlayarak e-ticaret duygu analizinde devrim yaratma potansiyelini göstermesidir.

Zümberoğlu ve diğ. (2025), mevcut veri setlerinin yetersiz boyutu, çeşitlilik eksikliği ve etiketleme tutarsızlıkları gibi sınırlamalar nedeniyle Türkçe duygu analizi alanındaki temel zorlukları ele almış ve bu problemleri çözmek amacıyla, otel, restoran, film, e-ticaret ürün incelemeleri ve tweetleri içeren çeşitli alanlardan toplanan verileri birleştiren, FSMTSAD adlı yeni ve geliştirilmiş bir veri seti sunmuştur. Bu veri seti, pozitif, nötr ve negatif olmak üzere üçlü sınıflandırma etiketlerine sahip 7.000 dengeli orijinal örnekten oluşmakta ve veri çeşitliliğini artırmak amacıyla geri çeviri ve eş anlamlı kelime değiştirme teknikleri kullanılarak toplam 15.853 örneğe genişletilmiştir. Çalışmada, hem geleneksel makine öğrenimi yaklaşımları LSVC, LR, NB, K-NN ve DT hem de transformer tabanlı modeller (ELECTRA Base Turkish, BERT Base Turkish, TurkishBERTweet, Turkish RoBERTa Base ve XLM-RoBERTa Base karşılaştırmalı olarak değerlendirilmiş; transformer modelleri için eğitim parametreleri (öğrenme oranı 1×10^{-5} , yığın boyutu 36) BounTi veri seti deneylerindeki ile aynı tutularak tutarlılık sağlanmıştır. Yapılan deneylerde en iyi performansı, FSMTSAD veri setinin test sonuçlarında %90,64 Doğruluk ve %90,64 F1 skoru elde eden ELECTRA Base Turkish modeli sergilemiştir. Makalenin temel

katkısı, Türkçe için dengeli, çeşitli ve büyük ölçekli yeni bir etiketli veri seti (FSMETSAD) sunması ve ELECTRA Base Turkish gibi modern transformer tabanlı modellerin bu geliştirilmiş veri setleri üzerinde önceki yaklaşımlara göre üstün performans gösterdiğini ampirik olarak kanıtlamasıdır



Çizelge 2.1 : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması

Çalışma	Veri Seti / Alanı	Veri Seti / Boyutu	Yöntemler	En İyi Yöntem	Sonuç
Lu ve diğ. (2010)	Konaklama - Çince	4.000	KTY: strength-based rules STY: adjective polarity/strength lexicon	RB	%71.65 kesinlik (2s), %38.88 kesinlik (5s)
Khan ve Baharudin (2011)	Sosyal Medya - İngilizce	1.500	KTY: sentence rules STY: SentiWordNet DDİY: POS ÖM: BoW	RB	%87 doğruluk
Jin ve diğ. (2012)	Çok Alanlı - Çince	982.166	ÖM: unigram, TF, TF-IDF, BOOL, CHI MÖ: SVM, BOOL-SVM, NB, KNN, OVA-SVMs, LR	OVA-SVM	%64.5 doğruluk
Vural ve diğ., (2013)	Film - Türkçe	60.000	KTY: sentence-binary, sentence-max/min STY: SentiStrength ÖM: word-sum	RB	%75.90 doğruluk
Vinodhini ve Chandrasekaran (2014)	E-ticaret - İngilizce	500	TY: Bagged SVM, Bayesian Boosting MÖ: SVM, LR	PCA+BB	%92.5 doğruluk

Çizelge 2.1 (devam) : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması

Çalışma	Veri Seti / Alanı	Veri Seti / Boyutu	Yöntemler	En İyi Yöntem	Sonuç
Xia ve diğ. (2015)	E-ticaret - İngilizce ve Çince	8.500	STY: DSA-WN, DSA-MI, LSS ÖM: unigrams; unigrams + bigrams MÖ: LR	SVM(2s), LR(3s)	%84.1 doğruluk (2s), %73.6 doğruluk (3s)
Vinodhini ve Chandrasekaran (2016)	E-ticaret - İngilizce	1.525	ÖM: N-grams (uni/bi/tri), TF-IDF MÖ: SVM TY: Bagging, Boosting, M-Bagging	M-Bagging	AUC: %89 G-ort: %90
Ciftci ve Apaydin (2018)	Çok Alanlı - Türkçe	355.095	ÖM: tf-idf, we MÖ: NB, LR DÖ: LSTM	LSTM (WE)	%82.9 doğruluk
Santur (2019)	E-ticaret - Türkçe	243.497	DÖ: GRU	GRU	%95.5 doğruluk, %97.6 F1
Aydın ve Güngör (2021)	Çok Alanlı - Türkçe ve İngilizce	21.960	MÖ: UL, SSL, SVM(tf-idf), SVM(3feats), MV(S), MV(U+S) DÖ: CNN, LSTM	MV (U+S)	%91.17 F1

Çizelge 2.1 (devam) : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması

Çalışma	Veri Seti / Alanı	Veri Seti / Boyutu	Yöntemler	En İyi Yöntem	Sonuç
Aktaş ve diğ. (2021)	Yemek - Türkçe	676.000	MÖ: SVM, kNN, NB, LR DÖ: NN	SVM	%87.24 doğruluk
Singh ve diğ. (2022)	Sosyal Medya - İngilizce	179.108	MÖ: NB, RF, SVM, LR DÖ: LSTM-RNN, FT-LSTM-RNN	FT-LSTM-RNN	%84.56 doğruluk, %81.23 F1
Ben ve diğ. (2023)	E-ticaret - İngilizce	36.000.000	MÖ: NB, RF, SVM DÖ: BiLSTM, LSTM, CNN-BiLSTM, CNN-LSTM, CNN	CNN	%85 doğruluk
Kumar ve diğ. (2023)	E-ticaret - İngilizce	50.000	DÖ: LSTM, Bi-LSTM TRF: BERT	BERT	%92.25 doğruluk
Alawi ve Bozkurt (2024)	Çok Alanlı - Türkçe	29.393	DÖ: CNN-BiLSTM, CNN-LSTM, BiLSTM-CNN, LSTM, CNN, BiLSTM	CNN-LSTM	%95.43 doğruluk, %95.43 F1

Çizelge 2.1 (devam) : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması

Çalışma	Veri Seti / Alanı	Veri Seti / Boyutu	Yöntemler	En İyi Yöntem	Sonuç
Latha ve Rao (2024)	E-ticaret - İngilizce	60.000	MÖ: SVM DÖ: ANN, LSTM, APSO-LSTM, MCNN	MCNN	%97.40 doğruluk, %81.20 F1
Özdemir ve diğ. (2024)	Konaklama - Türkçe	5.000	MÖ: LOGR, kNN, DVM, NB, DT, RF TRF: BERT-M, ELECTRA-Tr, ALBERT-Tr, BERT BDM: Turna	TURNA	%99.4 doğruluk, %99.10 F1
Çakıcı ve diğ. (2025)	Çok Alanlı - Türkçe	65.000	TRF: XLM-T, TSAM, TR-BERTweet, BERTurk	TSAM	%97 doğruluk, %97 F1
İncidelen ve Aydoğan (2025)	E-ticaret - Türkçe	150.000	TRF: XLM-RoBERTa, mBERT, BERTurk-32k, BERTurk-128k, ELECTRA-TR-S, ELECTRA-TR-B	BERTurk-32k	%83.69 doğruluk, %83.65 F1
Teke ve diğ. (2025)	E-ticaret - Türkçe	73.392	MÖ: SVM, RF, NB, LR, kNN, GB, DT TRF: mBERT, BERTurk, XLNet, DistilBERT	BERTurk	%96 doğruluk, %96 F1

Çizelge 2.1 (devam) : Duygu analizi ile ilgili literatür çalışmalarının sınıflandırılması

Çalışma	Veri Seti / Alanı	Veri Seti / Boyutu	Yöntemler	En İyi Yöntem	Sonuç
Zümberoğlu ve diğ. (2025)	Çok Alanlı - Türkçe	15.853	MÖ: L SVC, LR, NB, kNN, DT DÖ: ELECTRA-TR, BERTurk, Bert-ML, TurkishBERTweet, TR-RoBERTa, XLM-RoBERTa	ELECTRA-TR-Base	%90.64 doğruluk, %90.64 F1



3. METODOLOJİ

Bu bölüm, tez çalışmasının uygulama sürecinin bilimsel arkaplanını sunmak üzere hazırlanmıştır.

Öncelikle doğal dil işlemenin (DDİ) ve duygu analizinin (DA) temel kavramlarına kısaca değinildikten sonra duygu analizi yöntemleri detaylıca açıklanacaktır:

- Kural Tabanlı Yöntemler
- Makine Öğrenmesi ve Derin Öğrenme Yöntemler
- Transformer Tabanlı Modeller
- BDM / ÜYZ Yöntemleri

Çalışmanın önemli parçasını oluşturan BERT, BERTurk, ELECTRA, RoBERTa, DeBERTa ve ModernBERT gibi transformer tabanlı modeller ile duygu analizi yöntemlerine değinildikten sonra ince ayar ve optuna ile hiperparametre optimizasyonları ile model eğitimi gibi uygulama süreçlerinin temelleri açıklanmıştır. Son olarak doğruluk, makro-F1, kesinlik ve duyarlılık gibi performans metrikleri tanıtarak, modellerin değerlendirilmesinde kullanılan ölçütler özetlenmektedir.

3.1 Doğal Dil İşleme

Doğal Dil İşleme (DDİ), insan dilinin bilgisayarlar tarafından anlaşılması, analiz edilmesi ve üretilmesini amaçlayan yapay zekâ alanıdır ve metin, konuşma veya çok modlu dil verilerinin anlamlandırılmasıyla ilgili geniş bir metodolojik çerçeve sunar (Jurafsky ve Martin, 2023).

Geleneksel DDİ yöntemleri çoğunlukla kural tabanlı ve istatistiksel çıkarımlara dayanırken, modern DDİ yaklaşımı bağlamı modelleyebilen derin öğrenme ve transformer tabanlı yapılara yönelmiştir (Young ve diğ., 2018). Özellikle dikkat mekanizması ve iki yönlü bağlamsal temsil güçleri sayesinde transformer modelleri, dildeki uzun bağımlılıkları ve duygu ilişkileri önceki yöntemlere kıyasla çok daha başarılı bir şekilde yakalayabilmektedir (Vaswani ve diğ., 2017). Bu paradigma

değişimi, duygu analizi gibi bağlama duyarlı problemlerde performans artışı sağlamış, metinlerdeki ince kutuplaşma, ironik ifadeler ve çok anlamlı duygusal yapılar gibi zorlu dil fenomenlerinin daha doğru modellenmesini mümkün kılmıştır.

Metin sınıflandırma, makine çevirisi, konuşma tanıma, özetleme, soru-cevap sistemleri, duygu analizi, sohbet robotları, bilgi çıkarımı ve metin üretimi gibi pek çok uygulamada DDİ modelleri aktif olarak kullanılmaktadır (Jurafsky ve Martin, 2023). Buna rağmen DDİ hâlen çözülmesi gereken çok sayıda derin probleme sahiptir. Bunların başında anlam belirsizliği, çok anlamlılık ve bağlam bağımlılığı gelmektedir; zira kelimelerin anlamı çoğu zaman içinde buldukları bağlama göre değişmekte ve bu durum modellerin semantik çözümleme kapasitesini zorlamaktadır. Ayrıca ironi, sarkazm, mecaz, örtük duygu ve pragmatik çıkarım gibi insan diline özgü unsurların bilgisayar tarafından yakalanması, doğal dilin bilişsel yönlerinin modellenmesini gerektirdiği için hâlen zorlu bir araştırma konusudur (Wiebe ve diğ., 2005).

Öte yandan dijital platformlarda üretilen verilerin önemli bir kısmı gürültülü, hatalı, kısaltmalarla dolu, konuşma diline özgü veya birden fazla dilin karıştığı biçimdedir; bu da veri temizleme, normalizasyon ve model genellemesi açısından önemli teknik sorunlara yol açmaktadır. Diğer taraftan Türkçe gibi eklemeli dillere özgü bağlam-dildeğişimi (morfoloji) ve sözdizimi problemleri, hem kelime kökü tespiti hem de anlam çözümlemesi süreçlerinde modern modeller için ek zorluklar yaratmaktadır (Sak ve diğ., 2007).

Tüm bu zorluklara rağmen DDİ, sağlık hizmetlerinden finansal piyasalara, müşteri deneyiminden bilgi güvenliğine kadar geniş bir uygulama alanı sunmaktadır. Hastaların klinik notlarından otomatik tanı tahmini yapılabilmekte, bankacılık sektöründe dolandırıcılık tespiti için metin tabanlı anomali algılama sistemleri geliştirilebilmekte, e-ticarette müşteri memnuniyetini ölçmek için duygu analizi modelleri kullanılmakta, hukuki belgelerin otomatik sınıflandırılmasıyla karar destek sistemleri oluşturulabilmektedir. Yeni nesil büyük dil modelleri ise metin üretimi, çok dilli çeviri, bilgi çıkarımı, tutarlı diyalog yönetimi ve sıfır örnekle öğrenme gibi alanlarda DDİ'nin yeteneklerini genişleterek bu teknolojiyi stratejik bir inovasyon alanına dönüştürmüştür (Brown ve diğ., 2020). Dolayısıyla DDİ hem teknik hem bilişsel hem de toplumsal boyutlarıyla, modern yapay zekânın en geniş kapsamlı ve en hızlı gelişen araştırma alanlarından biri olmaya devam etmektedir.

3.2 Duygu Analizi

Duygu analizi, doğal dil işleme (DDİ) alanının en önemli alt başlıklarından biri olup, kullanıcıların metinler aracılığıyla ifade ettikleri duygu, tutum veya değerlendirmeleri otomatik olarak belirlemeyi amaçlayan bir sınıflandırma problemidir (Liu, 2012). Dijitalleşmeyle birlikte sosyal medya gönderileri, ürün yorumları ve müşteri geri bildirimleri gibi kullanıcı kaynaklı verilerin hızla artması, duygu analizini hem akademik literatürde hem de endüstriyel uygulamalarda kritik bir konuma getirmiştir. Güncel çalışmalar, duygu analizinin yalnızca olumlu–olumsuz sınıflandırmasından ibaret olmadığını; bağlamsal kutuplaşmanın tanınması, ironi ve sarkazm tespiti, duygu yoğunluğunun ölçülmesi ve çok boyutlu duygusal durumların modellenmesi gibi daha karmaşık problemlere doğru genişlediğini göstermektedir (Cambria ve diğ., 2017).

Duygu analizi yöntemleri tarihsel olarak üç ana başlık altında değerlendirilmektedir: kural tabanlı yöntemler, makine öğrenimi (MÖ) yöntemleri ve derin öğrenme (DÖ) yöntemleri. Kural tabanlı yöntemler, kelime sözlüklerine ve dilbilgisel göstergelere dayanarak duygu puanı hesaplayan en erken yaklaşımlardır (Taboada ve diğ., 2011). Bu yöntemler eğitim verisi gerektirmemeleri nedeniyle avantajlı olsa da çok anlamlılık, bağlamsallık ve ironi gibi karmaşık dil olgularını yeterince modelleyemezler (Wilson ve diğ., 2005). Makine öğrenimi yöntemleri bu sınırlılıkları aşmak için TF-IDF gibi sayısal metin temsilleri üzerinden sınıflandırma modelleri (ör. SVM, NB, KNN) kullanarak daha verimli bir yaklaşım sunmuştur (Pang ve diğ., 2002). Ancak bu yöntemler de kelimelerin bağlamsal ilişkilerini yakalama konusunda yetersiz kalmıştır.

Derin öğrenme yöntemlerinin ortaya çıkışıyla duygu analizi alanında önemli bir dönüşüm yaşanmıştır. Word2Vec (Mikolov ve diğ., 2013), GloVe (Pennington ve diğ., 2014) ve FastText (Bojanowski ve diğ., 2017) gibi kelime gömme modelleri sayesinde metinlerdeki semantik ilişkiler çok daha başarılı bir şekilde temsil edilmeye başlanmıştır. Ardından LSTM, GRU ve CNN gibi sinir ağı mimarileri bağlamsal bağımlılıkları yakalama kapasitesi sayesinde duygu analizinde klasik makine öğrenimi yöntemlerinin ötesine geçmiştir (Kim, 2014). Ancak asıl paradigma değişimi, dikkat mekanizmasını temel alan Transformer mimarisinin (Vaswani ve diğ., 2017) duygu analizi çalışmalarına entegre edilmesiyle gerçekleşmiştir. BERT (Devlin ve diğ., 2019), RoBERTa (Liu ve diğ., 2019), ELECTRA (Clark ve diğ., 2020) ve DeBERTa-

v3 (He ve diğ., 2021) gibi modeller, dilin bağlamsal yapısını iki yönlü ve derin temsil eden önceden eğitilmiş modeller aracılığıyla duygu analizinde sektör standardı hâline gelmiştir. Son yıllarda ise büyük dil modelleri (LLM) —ChatGPT, LLaMA, Claude, Gemma gibi— birkaç örnekle öğrenme, sıfır örnekle öğrenme ve talimatla yönlendirilmiş sınıflandırma gibi yetenekleri sayesinde duygu analizini eğitim gerektirmeyen veya minimum veriyle çalışan yeni bir seviyeye taşımıştır (Brown ve diğ., 2020).

3.3 Yöntem ve Modeller

Bu bölümde, tez çalışması kapsamında duygu analizi problemini ele almak amacıyla başvuru yöntem ve model yaklaşımları dört ana başlık altında sistematik bir biçimde özetlenmiştir. İlk aşamada, metin içerisinde yer alan kelime ve ifadelerin önceden tanımlanmış sözcüksel kaynaklar ve sözlük tabanlı puanlama mekanizmaları aracılığıyla değerlendirildiği kural tabanlı yöntemler ele alınmıştır. Ardından, duygu sınıflandırma problemlerinde yaygın olarak kullanılan geleneksel makine öğrenimi algoritmaları ile derin öğrenme temelli modellerin temel özellikleri ve uygulama biçimleri açıklanmıştır. Üçüncü alt başlıkta, güncel literatürde yüksek başarı oranlarıyla öne çıkan ve modern doğal dil işleme çalışmalarının temelini oluşturan transformer tabanlı modellerin mimari yapıları ve çalışma prensiplerine yer verilmiştir. Son olarak ise büyük dil modelleri ve üretken yapay zekâ yaklaşımlarının duygu analizi görevlerindeki konumu, kullanım alanları ve sağladıkları katkılar genel bir çerçeve içerisinde ele alınmıştır.

3.3.1 Kural tabanlı yöntemler

Duygu analizi alanında kural tabanlı yöntemler, metinlerin duygusal içeriğini (pozitif, negatif, nötr) belirlemek için önceden tanımlanmış dilbilimsel ve sözcüksel kaynakları kullanan klasik yaklaşımlardır. Bu yöntemler, metinlerin duygusal tonunu tayin etmek amacıyla kural setleri, kalıplar ve etiketlenmiş kelime listeleri kullanır. Makine öğrenimi tabanlı duygu analizi yaklaşımları, kural tabanlı yöntemlerden genellikle daha iyi performans gösterse de (Kolchyna ve diğ., 2015), kural tabanlı bileşenler, makine öğrenimi modelleri için girdi özelliği olarak kullanılarak sınıflandırma hassasiyetini artırabilmektedir (Mullen ve Collier, 2004). Ayrıca, dilsel kurallara

dayalı bu yapılar, alan bilgisinin doğrudan modele yansıtılmasına olanak tanıyarak hibrit duygu analizi yaklaşımlarının temelini oluşturmaktadır.

3.3.1.1 Duygu sözlük tabanlı yaklaşım

Duygu sözlüklerine dayalı puanlama, bir metindeki kelimeleri, önceden değer atanmış kelime listeleriyle eşleştirerek genel duygu skorunu hesaplayan temel kural tabanlı yaklaşımdır. Özellikle amaç, Twitter gibi platformlarında, metinlerin duygu gücünü hızlı ve basit bir şekilde tespit etmektir. Her kelimeye pozitif veya negatif bir puan atanan duygu sözlükleri ile skor hesaplanır.

Ayrıca, Twitter mesajları için kural tabanlı sınıflandırmada duygu sözlüklerinin emojiler, kısaltmalar ve sosyal medya argosu gibi unsurlarla zenginleştirilmesi de önerilmiş ve doğruluğu arttırdığı gösterilmiştir (Kolchyna ve diğ., 2015).

3.3.1.2 Bağlamsal kutuplaşma yaklaşımı

Kural tabanlı yaklaşımların en büyük zorluklarından biri, kelimenin duygusal değerini değiştiren bağlamsal etkileri ele almaktır. Bu tür dilbilimsel kurallar, cümlenin gerçek duygusal kutuplaşmasını doğru bir şekilde tanımak için hayati öneme sahiptir.

Bir kelimenin duygusal yönünün cümle içindeki bağlamına göre değişebilmesi önemli bir problem olarak görülür. Örneğin, "mükemmel" pozitifken, "mükemmel değil" negatif veya nötr hale gelir. Wilson ve diğerleri (2005) bu sorunu ele alarak, ifade düzeyinde bağlamsal kutuplaşmayı otomatik olarak tanımak için kural tabanlı bir yaklaşım sunmuşlardır. Bu yöntemler, bir kelimenin öznelliğini ve kutuplaşmasını otomatik olarak belirlemek için kullanılır. Daha sonraki çalışmalar, bu tür dilbilgisel yapıları, Naif Bayes gibi sınıflandırıcıların doğruluğunu önemli ölçüde artırarak, hızlı ve yüksek doğrulukta sınıflandırma yapılmasını sağlamıştır (Narayanan ve diğ., 2013).

3.3.2 Makine öğrenimi ve derin öğrenme yöntemleri

Makine öğrenmesi ve derin öğrenme yöntemleri, uzun yıllar boyunca duygu analizi alanında temel yaklaşım olarak kullanılmıştır. Bu yöntemlerde metinler önce sayısal bir forma dönüştürülür; örneğin kelime torbası (BOW), terim frekansı–ters belge sıklığı (TF-IDF) veya kelime gömme (WE) gibi teknikler kullanılır (Mikolov ve diğ., 2013). Bu temsiller daha sonra LR, SVM, RF ve NB gibi sınıflandırıcılarla işlenerek metnin pozitif, negatif veya nötr duygu taşıdığı tahmin edilir (Young ve diğ., 2018).

Bu modeller özellikle veri miktarı az olduğunda hızlı ve etkili çözümler sunar; ancak bağlamı yeterince anlayamadıkları için ironi ve alay gibi bağlama bağlı olguları doğru değerlendirmekte zorlanırlar (Cambria ve diğ., 2017).

Bu geleneksel yaklaşımları takip eden derin öğrenme modelleri duygu analizinde önemli bir ilerleme sağlamıştır. Word2Vec, GloVe ve FastText gibi kelime gömme yöntemleri kelimeler arası anlamsal ilişkileri daha doğru temsil ederken (Mikolov ve diğ., 2013), LSTM (Hochreiter ve Schmidhuber, 1997), GRU ve CNN gibi modeller metnin sırasal yapısını ve anlam örüntülerini öğrenerek daha yüksek doğruluk sunmuştur (Kim, 2014). Bu modellerden türetilen BiLSTM ve dikkat mekanizması, özellikle uzun yorumlarda önemli ifadelerin öne çıkarılmasını sağlayarak doğruluğu artırmıştır (Bahdanau ve diğ., 2015). Bununla birlikte, derin öğrenme yöntemleri yüksek hesaplama maliyeti ve büyük veri gereksinimi nedeniyle son yıllarda büyük ölçüde transformer tabanlı modellere bırakılmıştır (Vaswani ve diğ., 2017).

3.3.2.1 Tamamlayıcı naif bayes

Çoğul Naif Bayes (MNB) modelinin metin sınıflandırmasında yaygın olarak kullanılmasına rağmen, MNB'nin performansının düşmesine yol açan sistematik hataları bulunmaktadır. Bu hatalardan biri, eğitim verilerinin çarpıklığından kaynaklanan önyargıdır; yani, bir sınıfın diğerine göre daha fazla eğitim örneğine sahip olması durumunda karar sınırının ağırlıklarının taraflı hale gelmesidir. Bu önyargı, daha az eğitim verisine sahip sınıflar için ağırlıkların küçülmesine yol açar ve sınıflandırıcının bilmeden bir sınıfı diğerine tercih etmesine neden olur.

Bu sistematik hatayı gidermek amacıyla, sınıf dengeli Tamamlayıcı Naif Bayes (CNB) model önerilmiştir (Rennie ve diğ., 2003). CNB'de, geleneksel MNB'nin aksine, bir sınıf için parametreler tahmin edilirken o sınıfa ait olmayan tüm diğer sınıflardaki veriler kullanılır. Bu yaklaşımın temel amacı, her tahmin için daha dengeli miktarda eğitim verisi kullanmak ve böylece çarpık veri önyargısına karşı daha az hassas olan daha kararlı ağırlık tahminleri elde etmektir.

Tamamlayıcı Naif Bayes modelinde kullanılan yumuşatma parametresi (α), sıfır olasılık problemini önlemek ve nadir görülen terimlerin modele aşırı etki etmesini engellemek amacıyla kritik bir rol oynamaktadır. Hiperparametre olarak α değerinin uygun şekilde ayarlanması, modelin genelleme yeteneğini doğrudan etkileyerek, basit

olasılıksal varsayımlara sahip bu tür sınıflandırıcıların dahi performansının anlamlı ölçüde iyileştirilmesine olanak tanır.

3.3.2.2 K-en yakın komşuluk

K - En Yakın Komşuluk (K-NN), örüntü sınıflandırması alanındaki temel algoritmalarından biridir. K-NN, yeni bir veri noktasını sınıflandırmak için eğitim kümesindeki en yakın K komşusunun çoğunluk oyunu kullanır. Benzerlik genellikle Öklidyen mesafe, Manhattan mesafesi veya kosinüs benzerliği gibi mesafe ölçütleri aracılığıyla hesaplanır.

K-NN algoritması, bir örüntü sınıflandırıcısı olarak incelenmiştir (Duda ve Hart, 1973). Bu yaklaşımın performansı, temel olarak k değerinin seçimine ve kullanılan mesafe metriğine bağlıdır. Optimal k değerinin belirlenmesi, modelin genelleştirme yeteneği için kritik bir hiperparametre seçimi sorunudur. Küçük k değerleri modele yüksek duyarlılık (aşırı yakınsama) kazandırabilirken, büyük k değerleri sınıf sınırlarının aşırı düzleşmesine neden olabilir.

Ayrıca K-NN, tüm eğitim verisini bellekte tutması ve her tahmin için mesafe hesaplaması yapması nedeniyle hesaplama maliyeti yüksek bir yöntemdir. Buna karşın, model varsayımı içermemesi ve karmaşık veri dağılımlarında dahi uygulanabilir olması, K-NN'i özellikle karşılaştırma amaçlı temel bir algoritma olarak literatürde yaygın şekilde kullanılan bir yöntem hâline getirmektedir.

3.3.3 Transformer tabanlı modeller

Transformer mimarisi, Vaswani, Shazeer, Parmar ve diğer yazarlar (2017) tarafından dönüm noktası niteliği taşıyan "Attention Is All You Need" (İhtiyacınız Olan Tek Şey Dikkattir) başlıklı makale ile tanıtılmıştır. Bu mimarinin ele aldığı temel problem, o dönemde baskın olan sıralı çeviri modellerinin dayandığı karmaşık yinelemeli veya evrişimli sinir ağlarından kaynaklanan kısıtlamalardır (Vaswani ve diğ., 2017). Bu sıralı işlem, doğası gereği eğitim örnekleri içinde paralelleştirmeyi engelliyordu.

Transformer, bu problemleri çözmek için yineleme ve evrişimden tamamen vazgeçerek mimarisini yalnızca dikkat mekanizmalarına dayandırmayı önermiştir (Vaswani ve diğ., 2017). Model, Kodlayıcı-Kod Çözücü yapısını kullanır ve global bağımlılıkları modellemek için çok başlı öz-dikkat mekanizmasına güvenir. Bu yapı, hesaplamaların paralelleştirilmesine ve uzun menzilli bağımlılıkların daha etkin bir

şekilde modellenmesine olanak tanımıştır (Vaswani ve diğ., 2017). Böylelikle eğitim için belirgin ölçüde daha az zaman gerektirmesini sağlamıştır.

Bu modellerin başarısı, ön-eğitilmiş temsilleri ek özellikler olarak kullanan ELMo (Peters ve diğ., 2018) gibi özellik tabanlı yaklaşımlardan, ön-eğitilmiş parametrelerin görev bazında ince ayarlandığı Üretken Ön-Eğitilmiş Transformer (GPT) (Radford ve diğ., 2018) gibi ince ayar yaklaşımlarına doğru evrilen dil temsili öğrenimindeki köklü değişime dayanmaktadır (Rogers ve diğ., 2020). Ancak, ilk dil modelleri genellikle genel dil temsillerini öğrenmek için tek yönlü dil modelleri kullanma kısıtlamasını taşıyordu (Yang ve diğ., 2019). Bu sınırlamayı aşarak, BERT modelinin (Devlin ve diğ., 2019) piyasaya sürülmesiyle, DDİ'de çift yönlü ve derin öğrenilmiş temsiller dönemi başlamıştır.

BERT'in başarısından sonra, bu büyük ölçekli modellerin verimliliğini, sağlamlığını ve performansını artırmayı amaçlayan yoğun bir araştırma dalgası, yani BERToloji, ortaya çıkmıştır (Rogers ve diğ., 2020). Bu çalışmalar, ya daha sağlam optimizasyon stratejileri RoBERTa (Liu ve diğ., 2019), ya yeni ön-eğitim hedefleri XLNet (Yang ve diğ., 2019) ve ELECTRA (Clark ve diğ., 2020) ya da parametre verimliliğini artıran mimari değişiklikler ALBERT (Lan ve diğ., 2019) ve DeBERTa (He ve diğ., 2020) aracılığıyla orijinal modelin sınırlamalarını gidermeye odaklanmıştır.

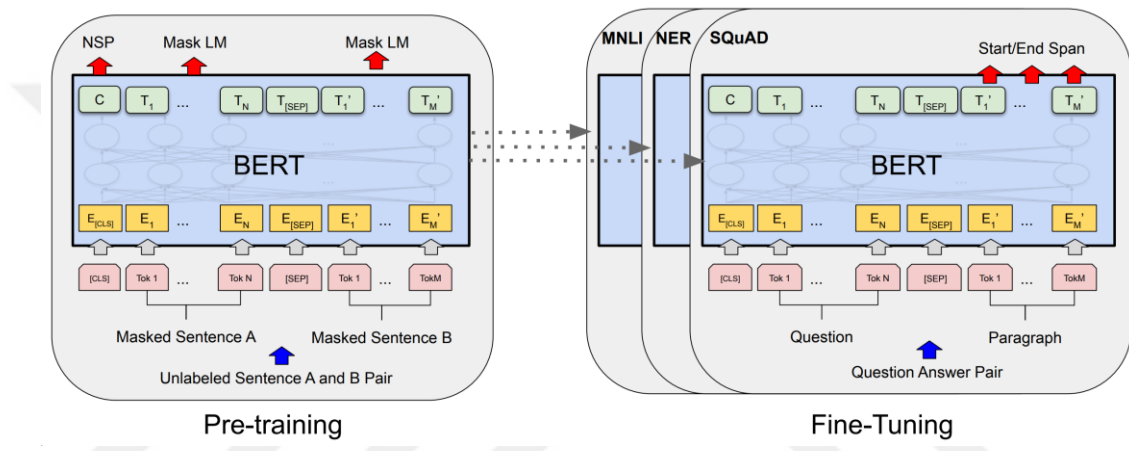
3.3.3.1 BERT

Transformer mimarisinin (Vaswani ve diğ., 2017) doğal dil işleme (DDİ) alanında kullanılmaya başlanması, Transformer'lardan Çift Yönlü Kodlayıcı Temsilleri (BERT) modeliyle (Devlin ve diğ., 2019) birlikte büyük bir dönüşüme yol açmıştır (Rogers ve diğ., 2020). Geleneksel olarak, dil modelleri ya kelime temsillerini bağlamdan bağımsız olarak öğreniyordu, ya da bağlamsallaştırılmış temsiller sunan ilk modeller bile, genel dil temsillerini öğrenmek için tek yönlü dil modellerini kullanma kısıtlamasını taşıyordu (Rogers ve diğ., 2020). Bu kısıtlama, bir token'ın temsilinin yalnızca solundaki veya sağındaki bağlama dayanmasına neden oluyordu. Bu durum, metin anlama, soru yanıtlama ve dil çıkarımı gibi görevlerde dilin karmaşık anlamsal özelliklerinin tam olarak modellenmesini zorlaştıran temel bir sorun teşkil ediyordu.

BERT modeli, bu sorunu çözmek için etiketsiz metinlerden derin çift yönlü temsillerin ön-eğitimi prensibini getirmiştir (Devlin ve diğ., 2019). BERT'in başarısı, tüm katmanlarda hem sol hem de sağ bağlama ortaklaşa koşullanması yeteneğine dayanır.

Bu çift yönlülüğü sağlamak için iki temel kendi kendine denetimli ön-eğitim görevi kullanılmıştır: Maskelenmiş Dil Modellemesi (MLM) ve Sonraki Cümle Tahmini (NSP). MLM, girdide rastgele maskelenen token'ların orijinal kimliğini tahmin etmeyi amaçlayarak modelin derin, çift yönlü bir Transformer kodlayıcısı (Vaswani ve diğ., 2017) eğitimini sağlar. NSP ise modelin metin çiftleri arasındaki ilişkiyi öğrenmesine yardımcı olan ikili bir sınıflandırma görevidir. BERT, bu yaklaşımla kavramsallık açısından basit olmasına rağmen ampirik olarak güçlü olduğunu kanıtlamış, GLUE ve SQuAD gibi karşılaştırmalı testlerde önemli performans artışları sağlamıştır.

Şekil 3.1'de BERT modelinin genel ön eğitim ve ince ayar süreçleri gösterilmektedir.



Şekil 3.1 : BERT ön eğitim ve ince ayar süreçleri (Devlin ve diğ., 2019)

Bu ön-eğitim stratejisi sayesinde; BERT, göreve özgü mimari modifikasyonlara gerek kalmadan, yalnızca tek bir çıktı katmanını eklenerek geniş bir görev yelpazesinde ince ayar yapılabilen esnek bir yapı sunmuştur (Rogers ve diğ., 2020).

3.3.3.2 RoBERTa

RoBERTa (Sağlamlaştırılmış ve Optimize Edilmiş BERT) çalışmasının ele aldığı temel problem, farklı ön-eğitim yaklaşımları arasındaki performans karşılaştırmalarının zorluğu ve orijinal BERT modelinin yeterince eğitilip eğitilmediği ile ilgiliydi (Liu ve diğ., 2019). Eğitim maliyetlerinin yüksek olması, farklı veri kümesi boyutlarının kullanılması ve hiperparametre seçimlerinin nihai sonuçlar üzerindeki önemli etkisi nedeniyle, BERT'ten sonra bildirilen performans iyileşmelerinin kaynağını kesin olarak belirlemek zorlaşmıştı (Liu ve diğ., 2019). Bu durumu açıklığa kavuşturmak için Liu ve diğerleri (2019), BERT ön-eğitimi üzerine dikkatli bir replikasyon çalışması yürütmüşlerdir.

RoBERTa'nın çözümü, BERT'in mimarisini koruyarak ön-eğitim sürecindeki temel tasarım kararlarını sağlam bir şekilde optimize etmek olmuştur (Liu ve diğ., 2019). Çalışmada, orijinal BERT'in önemli ölçüde az eğitilmiş olduğu tespit edilmiştir (Liu ve diğ., 2019). Çözüm, birden fazla optimizasyonun uygulanmasını içermiştir: Daha büyük toplu iş boyutları, daha uzun süre ve daha fazla veri üzerinde eğitim yapmak (Liu ve diğ., 2019); BERT'in metin çifti temsillerini öğrenmek için kullanılan NSP görevini tamamen kaldırmak (Liu ve diğ., 2019); ve eğitim verisine uygulanan maskeleme modelini statik yerine dinamik olarak değiştirmek (Liu ve diğ., 2019). Bu hiperparametre optimizasyonları ve eğitim veri boyutunun dikkatli ölçümü, RoBERTa'nın MLM ön-eğitim hedefinin etkinliğini yeniden kanıtlamasını sağlamıştır (Liu ve diğ., 2019). Elde edilen önemli sonuçlar, en iyi RoBERTa modelinin yayımlandıktan sonra çıkan XLNet (Yang ve diğ., 2019) dahil olmak üzere kendisinden sonra yayımlanan her modelin performansını yakaladığını veya geçtiğini göstermiştir (Liu ve diğ., 2019).

3.3.3.3 ELECTRA

ELECTRA (Verimli Yerine-Koyma Tabanlı Kodlayıcı), geleneksel MLM yerine daha örneklem-verimli bir ön-eğitim görevi olan Değiştirilen Token Tespiti (RTD) yaklaşımını önermiştir (Clark ve diğ., 2020).

Transformer tabanlı derin dil modelleri alanında, BERT gibi çığır açan yöntemler, girdinin bir kısmını [MASK] özel tokeni ile değiştirerek orijinal tokenleri tahmin etmeyi amaçlayan MLM görevini kullanmıştır (Clark ve diğ., 2020). Ancak bu yaklaşım, çıktının sadece maskelenen tokenlerin küçük bir alt kümesi (genellikle %15) üzerinden tahmin edilmesi nedeniyle yüksek hesaplama maliyeti gerektiriyordu (Clark ve diğ., 2020). Bu verimsizlik, özellikle büyük ölçekli modellerin eğitim maliyeti ve erişilebilirliği konusunda endişelere yol açmıştır (Clark ve diğ., 2020). Clark ve diğerleri (2020), bu problemi çözmek ve daha örneklem-verimli bir ön-eğitim sağlamak RTD yaklaşımını kullanan ELECTRA modelini önermiştir. RTD yaklaşımı, girdiyi maskelemek yerine, küçük bir üretici ağdan alınan makul alternatiflerle bazı tokenleri değiştirerek bozar (Clark ve diğ., 2020). Daha sonra, ayırmacı model, maskelenen tokenlerin kimliğini tahmin etmek yerine, girdideki her bir tokenin orijinal veri kümesinden mi geldiğini yoksa üretici tarafından mı değiştirildiğini ikili bir sınıflandırma ile tahmin etmeyi öğrenir (Clark ve diğ., 2020). Bu ayırmacı görev,

modelin yalnızca maskelenmiş %15'lik kısım yerine tüm girdi tokenlerinden öğrenmesini sağladığı için MLM'den çok daha verimli hale gelir (Clark ve diğ., 2020). Bu yeni ön-eğitim hedefi sayesinde ELECTRA, aynı hesaplama bütçesi altında BERT'ten önemli ölçüde daha iyi bağlamsal temsiller öğrenmiştir (Clark ve diğ., 2020). Elde edilen önemli sonuçlar, modelin küçük ölçekte dahi güçlü olduğunu göstermiştir; örneğin, tek bir GPU'da 4 günde eğitilen bir ELECTRA-Küçük modeli, GPT'den daha iyi performans göstermiştir (Clark ve diğ., 2020). Daha büyük ölçekte ise, ELECTRA-Büyük modeli, RoBERTa (Liu ve diğ., 2019) ve XLNet (Yang ve diğ., 2019) ile karşılaştırılabilir sonuçlar elde etmiş, ancak bu modellerin ön-eğitim için harcadığı hesaplama maliyetinin 1/4'ünden daha azını kullanmıştır (Clark ve diğ., 2020). RTD'nin verimliliği, ELECTRA'dan sonraki çalışmalarda, örneğin DeBERTaV3, da etkisini kanıtlamıştır (He ve diğ., 2021). Bu durum, ELECTRA'nın getirdiği RTD paradigmasının, Transformer tabanlı dil modellerinin daha hızlı ve daha verimli eğitilmesi yolunda kritik bir ilerleme olduğunu kanıtlamıştır.

3.3.3.4 DeBERTa

DeBERTa (Ayrıştırılmış Dikkat Mekanizmalı ve Çözümleme-Geliştirmeli BERT), başlangıçta BERT (Devlin ve diğ., 2019) ve RoBERTa (Liu ve diğ., 2019) modellerini iyileştirmek üzere Microsoft tarafından tasarlanmıştır. DeBERTa'nın çözdüğü temel problem, bağlamsal temsilin kalitesini artırmaktır. Bunu iki ana teknikle başarmıştır: ayrık dikkat mekanizması, içeriği ve pozisyonu ayrı vektörlerle kodlayarak kelimeler arası dikkat ağırlıklarının bu bağımsız temsiller üzerinden hesaplanmasını sağlamıştır; ve gelişmiş maske kod çözücü, mutlak pozisyon bilgisini maskelenmiş tokenleri tahmin etme katmanına dahil etmiştir (He ve diğ., 2020).

DeBERTaV3 modeli, original DeBERTa mimari üzerine inşa edilerek, MLM tabanlı ön-eğitim yöntemlerinin doğası gereği yüksek hesaplama maliyeti gerektirmesi ile örneklem verimsizliği problemini çözmeyi ve özellikle çok dilli ortamlarda performansı maksimize etmeyi amaçlamıştır (He ve diğ., 2021). DeBERTaV3, ilk olarak DeBERTa'da kullanılan MLM hedefini, daha örneklem-verimli bir ön-eğitim görevi olan RTD (Clark ve diğ., 2020) ile değiştirmiştir. Modelin verimliliğini daha da artırmak için, ELECTRA'daki gömme paylaşımının yaratabileceği ve model verimliliğini düşüren "ip çekme" dinamiklerini engellemek amacıyla Gradyan Ayrıştırılmış Gömme Paylaşımı (GDES) adında yeni bir yöntem sunulmuştur (He ve

diğ., 2021). Bu yeniliklerin çok dilli korpuslar üzerinde uygulanması sonucunda DeBERTaV3, çapraz dil anlama görevlerinde önemli başarılar elde etmiştir (He ve diğ., 2021). Elde edilen kayda değer sonuç, mDeBERTaV3 Temel modelinin XNLI görevinde %79.8 sıfır atış çapraz dil doğruluğu elde etmesidir; bu skor, XLM-R Temel model gibi güçlü çok dilli temellere kıyasla %3.6'lık belirgin bir iyileşme sağlamış ve modelin benzer yapıdaki çok dilli modeller arasında yeni SOTA (State-of-the-Art) oluşturmasını sağlamıştır (He ve diğ., 2021).

3.3.3.5 BERTurk

BERTurk, Türkçe dili için tasarlanmış, topluluk odaklı BERT modellerini ifade eder (Schweter, 2020). BERTurk projesi, BERT'in Transformer tabanlı mimarisini temel alarak, Türkçe'nin kendine özgü dilbilgisel ve morfolojik yapısına uygun yüksek performanslı bağlamsallaştırılmış dil temsilleri sağlamayı amaçlamıştır. Dilbilgisel yapıları ve anlamsal farklılıkları nedeniyle, genel amaçlı çok dilli modeller (mBERT gibi) yerine, Türkçe gibi diller için alan özgü veya dile özel modellerin eğitilmesi, genellikle daha yüksek performans sağlar (Schweter, 2020).

BERTurk modelleri, Türkçe'ye özel geniş bir veri kümesi üzerinde eğitilmiştir. Bu eğitim verileri, filtrelenmiş ve cümlelere ayrılmış Turkish OSCAR korpusunu, güncel bir Wikipedia dökümünü, çeşitli OPUS korpuslarını ve Kemal Oflazer tarafından sağlanan ek korpusları içermektedir. Bu korpusun toplam boyutu 35GB olup, yaklaşık 4.4 milyar token içermektedir (Schweter, 2020). Proje kapsamında, büyük/küçük harf duyarlı BERTurk modeli, daha büyük bir kelime dağarcığı boyutu (128k) kullanan modelleri ve BERT'in daha küçük, daha hızlı bir versiyonu olan DistilBERTurk gibi çeşitli varyantlar yayınlanmıştır (Schweter, 2020).

3.3.3.6 TurkishBERTweet

Transformer tabanlı büyük dil modellerinin DDİ alanındaki etkinliğinin kanıtlanmasından (Devlin ve diğ., 2019) sonra ortaya çıkan temel problem, bu genel modellerin alan özgü verilerin karmaşık yapısını yeterince iyi temsil edememesidir (Nguyen ve diğ., 2020).

Bu kapsamda BERTweet, BERT temel mimarisini temel alan ancak RoBERTa'nın ön-eğitim prosedürü (Liu ve diğ., 2019) kullanılarak eğitilen, İngilizce Tweetler için tasarlanmış ilk büyük ölçekli dil modelidir (Nguyen ve diğ., 2020). BERTweet, büyük

bir İngilizce Tweet korpusu üzerinde ön-eğitimden geçirilerek, Sözcük Türü Etiketleme (POS) ve Adlandırılmış Varlık Tanıma (NER) gibi DDİ görevlerinde RoBERTa ve XLM-R gibi güçlü modelleri geride bırakmıştır (Nguyen ve diğ., 2020). Bu, alan özgü veriler üzerinde ön-eğitimin, genel amaçlı modellere göre hedef alanda daha iyi performans sağladığını kanıtlamaktadır (Nguyen ve diğ., 2020).

Daha sonrasında ise TurkishBERTweet modeli, 900 milyon Tweet kullanılarak Türkçe sosyal medya metinleri için tasarlanmış ilk büyük ölçekli ön-eğitilmiş dil modeli sunulmuştur (Najafi ve Varol, 2023). Modelin mimarisi, BERTweet gibi BERT-temel model ile aynı mimariyi paylaşmakta, ancak Türkçe veri seti ile RoBERTa ön-eğitim prosedürü kullanılarak eğitilmiştir (Najafi ve Varol, 2023). TurkishBERTweet, aynı zamanda daha küçük girdi uzunluğu ile tasarlanarak, BERTurk gibi modellere göre daha hafif bir yapıya kavuşmuş ve önemli ölçüde daha düşük çıkarım süresi elde etmiştir (Najafi ve Varol, 2023). Modelin performansı, Duygu Sınıflandırması ve Nefret Söylemi Tespiti gibi iki kritik metin sınıflandırma görevi üzerinde değerlendirilmiştir (Najafi ve Varol, 2023). Elde edilen başarılı sonuçlar, TurkishBERTweet'in genellenebilirlik konusunda mevcut alternatifleri geride bıraktığını göstermiştir (Najafi ve Varol, 2023).

3.3.3.7 ModernBERT

BERT'in piyasaya sürülmesinden bu yana bu modellerde sınırlı sayıda Pareto iyileştirmesi gözlemlenmiştir. ModernBERT, bu Pareto iyileştirme eksikliğini gidermek ve sadece kodlayıcı modellerin hız, bellek verimliliği ve bağlam uzunluğu yeteneklerini artırmak amacıyla modern optimizasyonlar getirerek tasarlanan yeni bir dil temsil modelidir (Warner ve diğ., 2024).

ModernBERT, teknik ve ölçek açısından iddialı özellikler sunmaktadır. Model, 2 trilyon token üzerinde eğitilmiştir ve 8192'ye varan doğal dizi girdi uzunluğuna sahiptir. Bu uzun bağlam kapasitesi, özellikle tam dikkat mekanizmalarının dizi uzunluğuna göre karesel olarak artan bellek bağımlılığı kısıtlamasını aşan modeller için kritik bir iyileşmedir (Warner ve diğ., 2024). Model, hızlı, bellek açısından verimli olacak şekilde tasarlanmıştır ve özellikle yaygın kullanılan GPU'larda çıkarım için optimize edilmiştir (Warner ve diğ., 2024). ModernBERT'in temel avantajı, sadece güçlü sonuçlar sunmakla kalmayıp, aynı zamanda en hızlı ve en bellek verimli kodlayıcılardan olmasıdır (Warner ve diğ., 2024).

3.3.4 Büyük dil modelleri ve üretken yapay zeka

Büyük Dil Modelleri (BBM'ler), geniş veriler üzerinde, genellikle öz-denetimli öğrenme kullanılarak büyük ölçekte eğitilen ve geniş bir yelpazedeki alt görevlere adapte edilebilen yapay zekâ dil modelleridir (Bommasani ve diğ., 2021).

Üretken yapay zekâ (ÜYZ), insan tarafından oluşturulmuş geniş veri kümelerindeki örüntüleri ve ilişkileri makine öğrenimi modelleri aracılığıyla öğrenerek metin, görsel, ses, müzik ve video gibi yeni içerikler üretebilen yapay zekâ sistemlerini ifade eder (Law, 2024). ÜYZ, geleneksel yapay zekâ yaklaşımlarından farklı olarak yalnızca geçmiş veriler üzerinden tahmin yapmaya değil, büyük dil modelleri, görsel üretim modelleri ve çok modlu üretim araçları yoluyla tamamen yeni içerikler oluşturmaya odaklanır (Law, 2024). Özellikle ChatGPT gibi GPT mimarisine dayalı büyük dil modellerinin ortaya çıkışıyla, ÜYZ dünya çapında ilgi görmüş ve insan benzeri tutarlılıkta dil üretme, kullanıcı etkileşimlerinden öğrenme ve farklı alanlarda üretken çıktı üretme kapasitesiyle dikkat çekmiştir (Law, 2024).

BDM'ler, ikili veya üçlü duygu sınıflandırması gibi daha basit duygu analizi görevlerinde etkili çözümler sunabilmektedir (Zhang ve diğ., 2023). Bu modeller, sıfır-örnek ayarında bile ince ayarlanmış daha küçük dil modelleriyle (KDM'ler) rekabet edebilmekte veya onları geçebilmektedir (Zhang ve diğ., 2023). Duygu analizi görevini BERT gibi geleneksel sınıflandırıcılar yerine, genellikle metinden-metne dönüşüm veya istem tabanlı sınıflandırma yoluyla gerçekleştirir. İstem tabanlı sınıflandırmada iki temel yaklaşım öne çıkmaktadır. İlki, modele herhangi bir etiketli örnek verilmeden yalnızca görevin tanımıyla çıktı ürettiren Örneksiz Öğrenme yaklaşımıdır. İkincisi ise, modelin ağırlıkları güncellenmeden yalnızca birkaç örnek içeren bir istem üzerinden genelleme yapmasını sağlayan Az-Örnekli Öğrenme yöntemidir.

Büyük dil modelleri kullanılarak duygu analizi yapılan bir çalışmada (Zhang ve diğ., 2023) çeşitli bulgulara erişilmiştir:

- Kültürel Bağlam Kargaşası: BDM'lerin alay, ironi, mizah, ve kültüre özgü ifadeler gibi dildeki karmaşık nüansları anlama yeteneği hala zorluklar taşımaktadır. Bu tür nüanslar genellikle bağlama ve paylaşılan kültürel arka plan bilgisine bağlıdır (Zhang ve diğ., 2023).

- Gerçek Zamanlı Adaptasyon Zorluğu: Duygu ifadeleri sosyal medya gibi platformlarda sürekli geliştiğinden, modeller bu trendlere uyum sağlaması gerekir. Ancak mevcut BDM'lerin ince ayar veya yeniden eğitim esnekliğindeki sınırlamalar, bu hızlı dil evrimine ayak uydurmalarını kısıtlamaktadır (Zhang ve diğ., 2023).
- Büyük Boyut, Garantili Performans Değil: Daha büyük modellerin her zaman üstün performans garantilemediğini göstermektedir (Zhang ve diğ., 2023). Örneğin, Flan-UL2 (20B), GPT-3.5 serisi modellere (175B) kıyasla çok daha küçük olmasına rağmen benzer performans sergileyebilmiştir (Zhang ve diğ., 2023).

Üretken yapay zekâ, GPT serisi (Achiam ve diğ., 2023), PaLM serisi (Chowdhery ve diğ., 2023), LLaMA serisi (Touvron ve diğ., 2023) ve Gemini (Gemini Team, 2023) gibi çeşitli temel modeller etrafında gelişmektedir. Bu modeller genellikle yüksek parametre boyutlarında olup, yerel bilgisayar ortamında çalıştırılmaya uygun değildir. Bu kapsamda açık kaynak BDM'ler ve KDM'ler yayınlanmaya başlanmıştır. 4-10 milyar parametre aralığındaki quantize edilmiş modellerin yerel bilgisayarlarda düşük VRAM ile çalıştırılabilmesi, bu modelleri kısıtlı kaynak ile özel uygulamalar ve akademik çalışmalar için ideal hale getirmektedir.

Ollama, Büyük Dil Modellerinin (BBM) yerel olarak dağıtılması ve çalıştırılması için tasarlanmış, kurulum kolaylığı sunan popüler ve açık kaynaklı bir çalışma ortamıdır (Marcondes ve diğ., 2025). Bu kapsamda Çizelge 3.1 üzerinde 4-10b aralığındaki çeşitli modeller listelenmektedir (Ollama, 2025). Tez çalışması kapsamında ise Gemma3 ve llama3.2 modelleri kullanılmıştır.

Gemma modelleri, Google'ın Gemini araştırmalarından elde edilen teknolojiye dayanan açık kaynak dil modelleridir (Gemma Team, 2024). Llama model ailesi, Meta AI tarafından geliştirilen açık kaynak LLM'ler arasında önemli bir yere sahiptir (Llama Team, 2024). Bu modeller, özellikle pratik boyutlarda yüksek performans sunmak üzere tasarlanmıştır. Ayrıca, yerel ortamda çalıştırılabilmeleri sayesinde veri gizliliği, deney tekrarlanabilirliği ve donanım kaynaklarının doğrudan kontrolü gibi avantajlar sunarak akademik deneyler için uygun bir altyapı sağlamaktadır. Bu özellikler, tez kapsamında yürütülen duygu analizi deneylerinin kontrollü ve sürdürülebilir bir şekilde gerçekleştirilmesine olanak tanımıştır.

Çizelge 3.1 : Popüler Ollama büyük dil modelleri

Model Adı	Model Tipleri (Parametre)	İndirilme Sayısı
llama3.1	8b, 70b, 405b	101.900.000
deepseek-r1	düşünebilme, 1.5b, 7b, 8b, 14b, 32b, 70b, 671b,	61.200.000
mistral	7b	18.900.000
gemma3	270m, 1b, 4b, 12b, 27b	15.300.000
qwen2.5	0.5b, 1.5b, 3b, 7b, 14b, 32b, 72b	13.600.000
llama3	8b, 70b	10.800.000
llava	7b, 13b, 34b	9.400.000
phi3	3.8b, 14b	9.200.000
qwen3	düşünebilme, 0.6b, 1.7b, 4b, 8b, 14b, 30b, 32b, 235b	7.800.000
gemma2	2b, 9b, 27b	7.300.000
gemma	2b, 7b	5.300.000
qwen	0.5b, 1.8b, 4b, 7b, 14b, 32b, 72b, 110b	4.900.000
qwen2	0.5b, 1.5b, 7b, 72b	4.300.000
llama2	7b, 13b, 70b	4.000.000
codellama	7b, 13b, 34b, 70b	2.900.000
dolphin3	8b	2.700.000
olmo2	7b, 13b	2.600.000
llama2-uncensored	7b, 70b	1.300.000
orca-mini	3b, 7b, 13b, 70b	650.200
...

3.4 Model Eğitimi

Bu bölümde, ilgili modellerin duygu analizi görevi kapsamında eğitim süreçleri ile hiperparametre optimizasyonu süreçleri açıklanmaktadır.

Model eğitimi süreci, öncelikle ince ayar mimarilerinin değerlendirilmesiyle başlamaktadır; burada tam ince ayar, parametre-verimli ince ayar ve talimat-temelli ince ayar gibi farklı yöntemler açıklanmaktadır.

Ardından, model başarımını doğrudan etkileyen hiperparametre optimizasyonu süreçleri incelenmekte; klasik arama teknikleri ile daha gelişmiş bayes optimizasyonu yaklaşımlarının model doğruluğu, verimliliği ve genelleme yeteneğine katkıları tartışılmaktadır.

Bu kapsam dâhilinde hem ince ayar stratejileri hem de hiperparametre optimizasyonu, modelin öğrenme kapasitesini iyileştirmeye yönelik bütüncül bir çerçeve oluşturmaktadır.

3.4.1 İnce ayar mimarisi

İnce ayar (Fine-tuning), önceden eğitilmiş bir BDM veya başka bir DÖ modelini temel alarak, bu modeli daha küçük, belirli bir göreve veya domaine özgü etiketli bir veri kümesi üzerinde ek eğitimden geçirme sürecidir (Parthasarathy ve diğ., 2024). Bu yöntem, genellikle önceden eğitilmiş modelin parametrelerinin hedef görevin gereksinimlerine göre ayarlanmasını içerir (Parthasarathy ve diğ., 2024). Öğrenilen geniş ve genel amaçlı dil temsillerini korumayı amaçlarken, yeni görev için daha doğru sonuçlar üretmesine imkân kılar.

İnce ayar, sıfırdan model eğitmeye (ön-eğitim) kıyasla genellikle daha az etiketli veri gerektirir ve daha hızlı yakınsama sağlar, böylelikle hesaplama maliyetlerini önemli ölçüde düşürür (Parthasarathy ve diğ., 2024). İnce ayar yaklaşımı, özellikle BERT (Devlin ve diğ., 2018) ve RoBERTa (Liu ve diğ., 2019) gibi Transformer (Vaswani ve diğ., 2017) tabanlı dil modellerinin DDİ alanındaki başarısında kilit rol oynamıştır (Mosbach ve diğ., 2020).

Büyük Dil Modelleri için kapsamlı bir ince ayar süreci yedi aşamalı bir yaşam döngüsüne sahiptir: Veri Kümesi Hazırlama, Model Başlatma, Eğitim Ortamı

Kurulumu, İnce Ayar, Değerlendirme ve Doğrulama, Dağıtım ve İzleme/Bakım (Parthasarathy ve diğ., 2024)

İnce ayar yöntemleri, büyük ölçekli dil modellerinin belirli bir göreve veya kullanım senaryosuna uyarlanmasını sağlayan üç temel yaklaşım altında incelenmektedir. Bu yöntemlerin seçiminde; modelin parametre sayısı, hesaplama karmaşıklığı, mevcut donanım kapasitesi ve uygulamanın gerektirdiği esneklik gibi faktörler belirleyici olmaktadır. Tam İnce Ayar (TİA) yaklaşımı tüm model parametrelerinin güncellenmesini içerdiği için en yüksek uyarlanabilirliği sunarken, büyük modellerde maliyetli ve donanım açısından sınırlayıcı olabilmektedir. Parametre-Verimli İnce Ayar (PVİA) yöntemleri ise modelin yalnızca küçük bir bölümünü eğiterek aynı kaliteyi çok daha düşük maliyetle sağlamayı hedeflemekte, bu nedenle özellikle büyük dil modellerinin pratik uygulamalara entegrasyonunda yaygın şekilde tercih edilmektedir. Talimat-Temelli İnce Ayar ise üretken yapay zekâ modellerinin doğal dilde verilen talimatları daha doğru takip edebilmesi için kullanılan ve zero-shot / few-shot performansı doğrudan etkileyen bir tekniktir.

3.4.1.1 Tam ince ayar

Tam İnce Ayar (TİA), önceden eğitilmiş bir sinir ağı modelinin, hedeflenen yeni bir alt görev üzerinde eğitilirken, modelin başlangıçtaki tüm parametrelerinin ağırlıklarının güncellenmesini içeren transfer öğrenme metodolojisidir (Howard ve Ruder, 2018).

Bu yöntem, genellikle BERT gibi Transformer tabanlı modellerin, genel amaçlı ön-eğitim aşamasında edindiği zengin dil temsillerini, belirli bir görev için özelleştirmek amacıyla kullanılır. Literatürde bu yaklaşım genellikle "tüm modelin ince ayarı" ya da kısaca "tam ince ayar" olarak adlandırılır (Howard ve Ruder, 2018).

Tam İnce Ayar, özellikle BERT'in başarılı transfer stratejisi olarak kabul edilmiş ve endüstri standardı haline gelmiştir (Howard ve Ruder, 2018). Bu yaklaşım, önceden eğitilmiş modelin her katmanındaki her bir ağırlığın, yeni veriler ışığında değiştirilmesini gerektirir (Howard ve Ruder, 2018).

Tam ince ayar, küçük veya orta büyüklükteki veri setlerinde bile, ön eğitimden elde edilen genel bilgiyi koruma yeteneği nedeniyle güçlü performans sağlar (Howard ve Ruder, 2018). Ancak, tam ince ayarın önemli bir zorluğu bulunmaktadır:

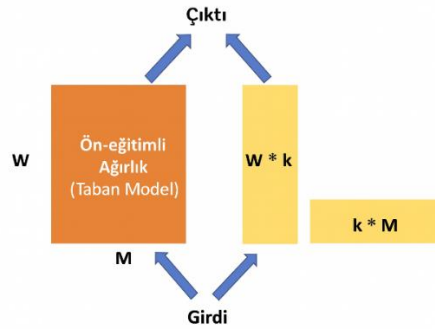
- **Maliyet ve Ölçeklenebilirlik:** Büyük Dil Modellerinin boyutu arttıkça, tüm parametreleri eğitmek çok pahalı hale gelir (Hu ve diğ., 2021). Her görev için tam ince ayar yapmak, önceden eğitilmiş modelin parametre sayısının N katı kadar parametre depolamayı ve eğitmeyi gerektirir (Houlsby ve diğ., 2019). Örneğin, LoRA gibi parametre-verimli yöntemler, GPT-3 175B gibi modellerde tam ince ayarlamaya kıyasla eğitilebilir parametre sayısını 10.000 kat azaltabilmektedir (Hu ve diğ., 2021). Modern araştırmalar çoğunlukla tam ince ayara benzer performans sağlayan, ancak çok daha az sayıda parametre güncelleyen (örneğin LoRA) parametre-verimli adaptasyon yöntemlerine odaklanmıştır (Hu ve diğ., 2021). Yine de, bazı çalışmalar, doğru hiperparametre optimizasyonu (Liu ve Wang, 2021) ve uygun optimizasyon teknikleri (Mosbach ve diğ., 2020) kullanıldığında tam ince ayarın istikrarının artırılabilirliğini göstermiştir.

3.4.1.2 Parametre verimli ince ayar

Parametre Verimli İnce Ayar (PVİA), büyük önceden eğitilmiş bir sinir ağı modelini belirli bir görev veya veri kümesine uyarlamak için, modelin parametrelerinin yalnızca çok küçük bir alt kümesinin güncellenmesini içeren bir dizi transfer öğrenme tekniğidir (Houlsby ve diğ., 2019)

Bu yöntem, Tam İnce Ayar yaklaşımının getirdiği ölçeklenebilirlik zorluklarına bir çözüm olarak ortaya çıkmıştır. Tam ince ayarın her görev için modelin tüm parametrelerini yeniden eğitme ihtiyacı, büyük dil modellerinde depolama ve hesaplama maliyetlerini aşırı derecede artırmaktadır.

Şekil 3.2’de görüldüğü üzere, PVİA yöntemleri; önceden eğitilmiş modelin parametrelerinin büyük bir çoğunluğunu dondurarak (sabit tutarak), yalnızca sınırlı sayıda (genellikle eklenen) parametreyi günceller (Hu ve diğ., 2021).



Şekil 3.2 : Parametre verimli ince ayar (George, 2024)

PVIA yaklaşımının temel faydaları şunlardır:

1. Maliyet Verimliliği: Eğitim için gereken GPU bellek ve hesaplama maliyetini düşürür (Parthasarathy ve diğ., 2024).
2. Depolama Verimliliği: Her yeni görev için tüm model ağırlıklarını depolamak yerine, yalnızca göreve özgü küçük bir parametre kümesini depolamayı gerektirir (Hu ve diğ., 2021).
3. Aşırı Unutmayı Engelleme: Önceden eğitilmiş temel model parametrelerini dondurarak, modelin yeni verilerle eğitilirken eski, genel bilgisini kaybetme-unutma sorununu hafifletmeye yardımcı olur (Parthasarathy ve diğ., 2024).
4. Performans: PEFT yöntemleri, çoğu zaman tam ince ayarlamaya benzer veya düşük veri hacimli senaryolarda daha iyi performans sunar (Parthasarathy ve diğ., 2024).

PVIA genellikle büyük dil modellerinin ince ayarı sırasında kullanımı tercih edilmekte olup çeşitli yöntemleri mevcuttur. Bu yöntemlerden Adaptörler, temel modeli dondurarak Transformer katmanları arasına yeni, eğitilebilir modüller ekler ve tam ince ayara yakın performans gösterir (Houlsby ve diğ., 2019). Buna karşılık LoRA (Hu ve diğ., 2021), ağırlık güncellemelerini düşük ranklı matrisler ile modelleyerek parametre sayısını 10.000 kata kadar azaltabilir ve çıkarım sırasında ek gecikme yaratmaz (Hu ve diğ., 2021). QLoRA (Dettmers ve diğ., 2023) ise, LoRA'nın bir uzantısı olarak, temel modeli 4-bit niceleyerek bellek verimliliğini önemli ölçüde artırır, bu da 65 milyar parametrelili modellerin tek bir GPU üzerinde ince ayar yapılmasına olanak tanır (Parthasarathy ve diğ., 2024).

3.4.1.3 Talimat temelli ince ayar

Talimatla-Temelli İnce Ayar (TTIA), önceden eğitilmiş bir Büyük Dil Modelini, doğal dil talimatları şeklinde ifade edilen bir dizi veri kümesi üzerinde ek eğitimden geçirme sürecidir (Wei ve diğ., 2022). Bu yöntem, modelin kullanıcının niyetini daha iyi anlamasını ve verilen talimatlara uygun yanıtlar üretme yeteneğini (zero-shot) geliştirmeyi amaçlar (Wei ve diğ., 2022).

Örneksiz Öğrenme, önceden eğitilmiş bir modelin, o görev için kendisine o göreve ait etiketli örnekler sunulmadan, bir görevi genelleme ve gerçekleştirme yeteneği olarak tanımlanır (Wei ve diğ., 2022). TTIA'nın en önemli çıktılarından biri de modellerin bu "örneksiz genelleme" kapasitesini artırmasıdır (Wei ve diğ., 2022). Az-Örnekli

Öğrenme ise modelin ağırlıklarını güncellemeden, yalnızca çok sınırlı sayıda etiketli eğitim örneği içeren bir girdi-istem sunularak yeni bir görevi gerçekleştirmesini sağlama yöntemidir (Brown ve diğ., 2020).

3.4.2 Hiperparametre optimizasyonu

Makine öğrenimi (MÖ) sistemlerinde merkezi bir rol oynayan hiperparametreler (HP), makine öğrenimi modellerinin karmaşıklığını kontrol etmek için önceden ayarlanması gereken konfigürasyon ayarlarıdır (Pedregosa, 2016). Model performansının etkinliği, büyük ölçüde hiperparametrelerin doğru seçimine bağlıdır (Shekhar ve diğ., 2022).

Hiperparametre Optimizasyonu (HPO), hiperparametre değerlerinin sistematik biçimde seçilmesini amaçlayan bir süreçtir. Bu zahmetli ve zaman alıcı süreci otomatikleştirmeyi hedefler. Hiperparametrelerin uygun bir şekilde ayarlanması hem model doğruluğu açısından kritik hem de hesaplama açısından zorlayıcıdır (Pedregosa, 2016). HPO, genel olarak, değerlendirilmesi pahalı ve gürültülü bir amaç fonksiyonunu en az sayıda değerlendirme ile minimize etme problemidir (Gardner ve diğ., 2014). Modelin performansını ve genelleme yeteneğini belirleyen bu kritik süreçteki bazı önemli hiperparametreler aşağıdaki gibidir:

- **Öğrenme Oranı (Learning Rate):** Öğrenme hızı, model ağırlıklarının her bir iterasyonda ne kadar güncelleneceğini belirleyen sürekli değerli bir hiperparametredir (Franceschi ve diğ., 2024). Modelin öğrenme sürecindeki her bir adımda ağırlıkları ne kadar değiştireceğini belirleyen temel hiperparametredir. Çok yüksek bir değer öğrenmeyi dengesiz hâle getirirken, çok düşük bir değer eğitimi yavaşlatır veya yerel minimumlara sıkışmasına neden olabilir.
- **Epoch Sayısı (Number of epochs):** ise modelin tüm eğitim verisini baştan sona kaç kez gördüğünü ifade eder. Daha fazla epoch genellikle daha fazla öğrenme anlamına gelse de, aşırı yüksek epoch değerleri overfitting riskini artırabilir.
- **Ağırlık Azalışı (Weight Decay):** modeldeki ağırlık parametrelerinin gereksiz yere büyümesini engellemek için kayıp fonksiyonuna eklenen bir düzenleme terimidir. Aşırı öğrenmeyi azaltmak amacıyla parametre büyüklükleri cezalandırılır (Ng, 2004).

- Isınma Oranı (Warmup ratio): öğrenme oranının eğitim başlangıcında öğrenme hızının kademeli olarak düşük bir değerden belirlenen hedef değere kadar kademeli olarak arttırılmasını için kullanılan bir parametredir. Transformer tabanlı modellerde, eğitim sürecinin başında ani gradyan değişimlerini önlemek amacıyla önemli bir stabilizasyon aracı olarak kullanılır (Vaswani ve diğ., 2017).
- Yığın Boyutu (Batch Size): modelin bir güncelleme adımında aynı anda işlediği örnek sayısını ifade eder. Küçük yığın boyutları daha gürültülü ama genellemesi kuvvetli güncellemeler sağlarken, büyük yığın boyutları daha stabil ancak daha yüksek bellek gerektiren optimizasyon süreçleri oluşturur (Masters ve Luschei, 2018). Büyük değer seçilmesi ile Ekran kartı bellek VRAM kapasitesi bu noktada sınırlayıcıdır. Eğitim ve doğrulama adımları için ayrı ayrı tanımlanabilir.
- Gradyan Biriktirme Adımları (Gradient accumulation steps): büyük mini parti büyüklüğü kullanımını GPU bellek sınırlamalarını aşmadan mümkün kılmak için kullanılan bir tekniktir. Bu yöntemde model gradyanları birkaç adım boyunca biriktirir ve yalnızca belirli bir adım sayısından sonra güncellemeyi gerçekleştirir (Micikevicius ve diğ., 2018). Böylelikle eğitim süresi minimize edilebilir.

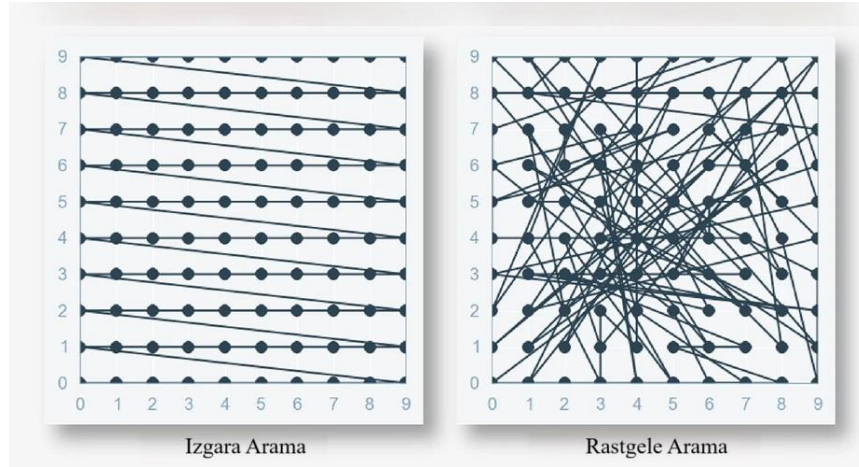
Bu temel hiperparametrelerin yanı sıra, dropout oranı, maksimum dizi uzunluğu, gizli katman sayısı, dikkat başlıklarının sayısı ve gradyan sınırlandırma gibi ek parametreler de modelin kararlılığı ve performansı üzerinde belirleyici rol oynar.

Bu değişkenlerin ayarlanması, sistemlerin nihai etkinliğini belirlediği için kritik öneme sahiptir (Franceschi ve diğ., 2024). Bir tekniğin performansının algoritmanın kendisinden mi yoksa ince ayarından mı kaynaklandığını anlamayı zorlaştırabilir (Bergstra ve diğ., 2013). Bu sebeple bu değişkenlere, hangi değer atamalarının yapılması gerektiği hususunda çeşitli yöntemler vardır, ve doğru yöntemin kullanılması önem arz eder.

3.4.2.1 Klasik optimizasyon

Şekil 3.3'de görüldüğü üzere rastgele arama (random search) ve ızgara arama (grid search) gibi klasik yöntemler geniş arama uzaylarında maliyetli ve verimsiz

olabileceğinden, bu yöntemlerin özellikle yüksek boyutlu derin öğrenme modelleri için yetersiz kaldığı sıkça vurgulanmaktadır (Bergstra ve Bengio, 2012).



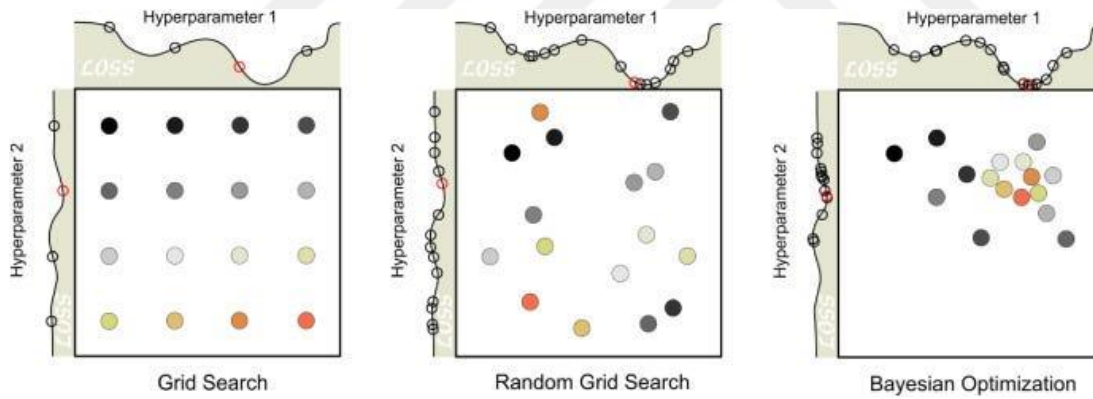
Şekil 3.3 : Klasik optimizasyon yöntemleri (Behera, 2024)

- **Izgara Arama (Grid Search):** önceden belirlenen tüm hiperparametre aralık kombinasyonlarını sistematik biçimde tarayan deterministik bir yaklaşımdır ve özellikle erken dönem makine öğrenimi çalışmalarında standart yöntem olarak kabul edilmiştir (Bergstra ve Bengio, 2012). Ancak hiperparametre uzayının boyutu arttıkça ızgara aramanın hesaplama maliyeti üstel biçimde büyümekte, bu da yöntemi yüksek boyutlu modeller için **verimsiz** hâle getirmektedir (Bergstra ve Bengio, 2012). Sayısal ve tamsayı hiperparametre değerleri tipik olarak **eşit aralıklarla** düzenlenir ki öğrenme hızı gibi bazı sürekli hiperparametrelerin, logaritmik bir ölçekte ayrıklaştırılması genellikle daha faydalı olmaktadır (Franceschi ve diğ., 2024). Model performansı üzerindeki etkiye sahip olan hiperparametreler, mevcut tüm hiperparametrelerin yalnızca küçük bir alt kümesidir (Bischl ve diğ., 2023). Izgara arama, bu önemsiz boyutlarda (etkisi olmayan hiperparametreler) zaman harcadığı için verimsiz olduğu ifade edilmektedir. Ayrıca gereken değerlendirme sayısı, hiperparametre sayısı ile üstel olarak artması da bir diğer verimsizlik noktasıdır.
- **Rastgele Arama (random search):** hiperparametre alanından rastgele örneklem alarak model performansını optimize etmeyi amaçlayan istatistiksel temelli bir yöntemdir ve yüksek boyutlu parametre uzaylarında ızgara aramadan çok daha etkili olduğu gösterilmiştir (Bergstra ve Bengio, 2012). Bağımsız örnekleme, esneklik ve paralelleştirme gibi avantajları olmasına

karşın, yöntemin rastlantısal yapısı sebebiyle en iyi hiperparametre kombinasyonunun garanti edilmemesi ve bazı kritik bölgelerin rastgelelik nedeniyle hiç ziyaret edilmemesi de dezavantajlar arasında gösterilmektedir. Yine de literatür, maliyet ve performans dengesi gözetildiğinde rastgele arama'nın özellikle geniş parametre uzaylarında ızgara aramaya göre daha uygulanabilir ve sonuç odaklı bir yöntem sağladığını göstermektedir (Bergstra ve Bengio, 2012).

3.4.2.2 Bayes optimizasyon

Son yıllarda hiperparametre optimizasyonunda daha verimli yaklaşımlar geliştirilmiştir. Geleneksel yöntemlerin (ızgara veya rastgele arama) endüstriyel ölçekli uygulamalarda zorluklar yaratması nedeniyle, Bayes Optimizasyon (BO) gibi daha gelişmiş stratejiler popülerlik kazanmıştır. Şekil 3.4'de görüldüğü üzere geleneksel yöntemler en iyi hiperparametre uzayının biraz ötesinde gereksiz aramalara da odaklanmaktayken, bayes optimizasyonu ise en iyi parametre aralıklarına daha çok önem vermektedir.



Şekil 3.4 : Klasik ve bayes optimizasyon karşılaştırma (Passos ve diğ., 2022)

BO, arama sürecini geçmiş gözlemlerden öğrenerek bir sonraki en iyi denemenin olasılıksal olarak tahmin edilmesini sağlar ve bu nedenle değerlendirilmesi zaman alan ve yüksek hesaplama maliyetli modellerde güçlü bir avantaj sunar (Mockus, 1975; Snoek ve diğ., 2012). BO, makine öğrenimi algoritmalarının hiperparametre ayarı ve deneysel tasarım gibi çeşitli problemlere başarıyla uygulanmıştır. Ayrıca Bayes optimizasyonunun yüksek boyutlu hiperparametre alanlarında klasik ızgara araması veya rastgele arama yöntemlerine kıyasla çok daha az deneme ile yüksek performanslı çözümler bulduğu, sayısal performans karşılaştırmalarıyla tekrarlı biçimde kanıtlanmıştır (Brochu, Cora ve de Freitas, 2010).

Bayes optimizasyonunun uygulanmasında kullanılan çok sayıda yöntem, araç ve yazılım bulunmaktadır. En yaygın yaklaşımlar arasında Gaussian süreçleri, beklenen iyileşme ve bilgi temelli edinim fonksiyonları gibi metodolojiler yer almaktadır (Snoek ve diğ., 2012). Uygulama tarafında ise Optuna, HyperOpt, SMAC3 ve BoTorch gibi kütüphaneler, Bayes optimizasyonunu farklı arayüzler ve optimizasyon stratejileri ile destekleyerek araştırmacılara esnek bir çalışma ortamı sunmaktadır. Bu araçlar, özellikle derin öğrenme modellerinin hiperparametre ayarlama süreçlerinde hata oranlarını düşürmek ve hesaplamayı hızlandırmak amacıyla yaygın bir şekilde kullanılmaktadır.

2018 yılında yayınlanan makalede, hiperparametre optimizasyon kütüphaneleri olan HyperOpt, Optuna, Optunity ve SMAC'i iki farklı benchmark üzerinde karşılaştırma yapılarak, çalışma süresi ile performans puanı arasındaki denge göz önüne alındığında Optuna'nın en uygun seçenek olduğu görüldüğü ifade edilmiştir (Shekhar ve diğ., 2022).

Optuna, hem bayes optimizasyonu hem de erken durdurma stratejilerini birleştirerek dinamik, verimli ve uyarlanabilir arama algoritmaları sunar (Akiba ve diğ., 2019). Böylelikle, büyük derin öğrenme modellerinde gereksiz eğitim döngülerini ortadan kaldırarak hesaplama maliyetini önemli ölçüde azaltarak optimum bir performansı garantiler.

Optuna'nın önerdiği temel tasarım kriterleri şunlardır (Akiba ve diğ., 2019):

1. Tanımla-ve-Çalıştır (Define-by-Run) API: Kullanıcıların parametre arama uzayını dinamik olarak oluşturmasına olanak tanıyan bir API sunar (Akiba ve diğ., 2019). Optuna, bu ilkeyle tasarlanmış ilk optimizasyon yazılımıdır.
2. Arama ve Budama (Pruning) Stratejilerinin Verimli Uygulanması: Hem arama hem de budama stratejilerinin verimli bir şekilde uygulanmasını hedefler. Böylelikle hedef fonksiyonu doğrultusunda düşük değerli denemeler önceden kesilerek, performans artışı sağlanır.
3. Kolay Kurulum ve Çok Yönlü Mimari: Ölçeklenebilir dağıtılmış bilgi işlemden etkileşimli arayüz aracılığıyla yürütülen hafif deneylere kadar çeşitli amaçlar için kolayca kurulabilen ve dağıtılabilen çok yönlü bir mimariye sahiptir.

Bu gelişmeler, duygu analizi gibi özellikle büyük veri ve karmaşık model yapılarının bulunduğu alanlarda hiperparametre seçiminin doğruluğunu artırarak daha güçlü sonuçlar elde edilmesine olanak tanımaktadır.

3.5 Performans Metrikleri

Duygu analizi, yazılı metinlerde ifade edilen öznel tutumları, duyguları veya değerlendirmeleri pozitif, negatif veya nötr gibi anlamsal yönelimlere otomatik olarak kategorize etmeyi amaçlayan temel bir doğal dil işleme görevi olmakla birlikte; bu analiz sonuçlarının ne kadar başarılı olduğunu belirlemek için performans değerlendirme metriklerine ihtiyaç duyulmaktadır (Choi ve Lee, 2017).

Bu metrikler, bir sınıflandırıcının etkinliğini ölçmek ve modelin güvenilirliğini doğrulamak için kritik öneme sahiptir (Vinodhini ve Chandrasekaran, 2014) ve temel olarak Genel Metrikler ve Sınıf Bazlı Metrikler olmak üzere iki sınıfta incelenebilir.

Bu bağlamda, sınıf bazlı metrikler her bir duygu sınıfı için modelin ayırt edici performansını ayrıntılı olarak incelemeye olanak tanırken, genel metrikler tüm sınıflar üzerindeki ortalama başarının daha bütüncül bir değerlendirmesini sunmaktadır. Özellikle sınıf dağılımının dengesiz olduğu duygu analizi problemlerinde, yalnızca doğruluk metriğine dayalı değerlendirmelerin yanıltıcı olabileceği göz önünde bulundurulurken, kesinlik, duyarlılık ve F1 skoru gibi tamamlayıcı metriklerin birlikte ele alınması gerekmektedir. Bu nedenle, bu bölümde kullanılan performans metrikleri hem sınıf bazında hem de mikro, makro ve ağırlıklı ortalama yaklaşımları çerçevesinde detaylı biçimde açıklanarak, tez kapsamında geliştirilen modellerin karşılaştırmalı ve adil bir şekilde değerlendirilmesi amaçlanmaktadır.

3.5.1 Sınıf bazlı metrikler

Sınıf bazlı metriklerden kasıt, modelin pozitif, negatif ve nötr gibi belirli bir sınıf üzerindeki performansını detaylı olarak incelemek üzerine belirlenen metrikleri temsil eder.

Sınıflandırma performansının hesaplanmasında kullanılan tüm metrikler, modelin tahminleri ile gerçek etiketler arasındaki ilişkiyi gösteren Karmaşıklık Matrisi (Confusion Matrix) bileşenlerine dayanmaktadır (Singh ve diğ., 2022). Karmaşıklık matrisinde kullanılan temel metrikler Çizelge 3.2’de gösterilmektedir.

Çizelge 3.2 : Karmaşıklık matrisi

Temel Metrik	Örnek
TP: Doğru Pozitif	Yorum pozitif → Model pozitif dedi.
TN: Doğru Negatif	Yorum negatif → Model negatif dedi.
FP: Yanlış Pozitif	Yorum negatif → Model pozitif dedi.
FN: Yanlış Negatif	Yorum pozitif → Model negatif dedi.

Bu temel bileşenler şunlardır (Aktaş ve diğ., 2021):

- Doğru Pozitif (TP): Modelin pozitif olarak tahmin ettiği ve gerçekte de pozitif olan örnek sayısıdır.
- Doğru Negatif (TN): Modelin negatif olarak tahmin ettiği ve gerçekte de negatif olan örnek sayısıdır.
- Yanlış Pozitif (FP): Modelin pozitif olarak tahmin ettiği, ancak gerçekte negatif olan örnek sayısıdır.
- Yanlış Negatif (FN): Modelin negatif olarak tahmin ettiği, ancak gerçekte pozitif olan örnek sayısıdır.

Bu temel bileşenler, sınıflandırma modellerinin farklı hata türlerini ayrıntılı biçimde incelemeye olanak tanımakta ve performans metriklerinin matematiksel olarak tanımlanmasının temelini oluşturmaktadır. Bu kapsamda doğruluk, kesinlik, duyarlılık, özgüllük ve F1 skoru gibi ölçütler, söz konusu bileşenler arasındaki ilişkiler üzerinden türetilerek model başarımının çok boyutlu bir şekilde değerlendirilmesini mümkün kılmaktadır.

3.5.1.1 Doğruluk

Denklem (3.1) üzerinden doğruluk hesaplamasının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

Doğruluk (Accuracy), en yaygın kullanılan performans metriğidir (Singh ve diğ., 2022). Kavramsal olarak doğruluk, bir modelin tahminde ne kadar başarılı olduğunun genel bir ölçüsünü sunsa da, özellikle sınıflar arasında dengesizlik bulunan veri setlerinde, baskın olan sınıfı her zaman doğru tahmin ederek yüksek yanıltıcı sonuçlar

verebilir (Chawla ve diğ., 2002). Bu nedenle, doğruluğun tek başına bir performans göstergesi olarak kullanılması yeterli değildir (Singh ve diğ., 2022).

3.5.1.2 Kesinlik

Denklem (3.2) üzerinden kesinlik hesaplamasının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (3.2)$$

Kesinlik (Precision), modelin pozitif olarak sınıflandırdığı örnekler arasında ne kadarının gerçekten pozitif olduğunu gösterir (Kanakamedala ve diğ., 2023). Yanlış pozitif tahminlerin (FP) maliyetinin yüksek olduğu durumlarda önemli bir metriktir (Singh ve diğ., 2022). Örneğin, bir spam filtresinde geçerli bir e-postayı yanlışlıkla spam olarak işaretlemek.

3.5.1.3 Duyarlılık

Denklem (3.3) üzerinden duyarlılık hesaplamasının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (3.3)$$

Duyarlılık (Recall), gerçekte pozitif olan tüm örnekler arasından modelin ne kadarını doğru bir şekilde pozitif olarak belirlediğini gösterir (Powers, 2011). Yanlış negatif tahminlerin (FN) maliyetinin yüksek olduğu durumlarda kritik öneme sahiptir (Singh ve diğ., 2022). Örneğin bir tıbbi testte hasta bir kişiye yanlışlıkla sağlıklı demek gibi.

3.5.1.4 Özgüllük

Denklem (3.4) üzerinden özgüllük hesaplamasının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Özgüllük} = \frac{TN}{TN + FP} \quad (3.4)$$

Özgüllük (Specificity), bir sınıflandırma modelinin, gerçekte negatif olan vakaları ne kadar doğru bir şekilde negatif olarak tanımladığını gösteren bir performans ölçütüdür (Powers, 2011). Bu metrik, genellikle Ters Duyarlılık (Inverse Recall) olarak da

adlandırılır (Powers, 2011). Örnek olarak, bir siber güvenlik sisteminde, yanlış alarm oranını düşürerek güvenlik ekibinin yalnızca gerçekten şüpheli aktiviteleri incelemesini mümkün kılmak üzerine özgüllük metriği ön plana çıkmaktadır.

3.5.1.5 F1 Skoru

Denklem (3.5) üzerinden fl skoru hesaplamasının nasıl gerçekleştirildiği gösterilmektedir.

$$F1 = \frac{2 \cdot \text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.5)$$

F1 Skoru, Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasıdır (Singh ve diğ., 2022). Bu metrik, hem kesinliği hem de duyarlılığı dengeli bir şekilde değerlendiren tek bir skor sunarak, özellikle sınıflar arası dağılımın dengesiz olduğu veri setlerinde performansı ölçmek için sıklıkla tercih edilir (Kanakamedala ve diğ., 2023),

3.5.2 Genel metrikler

Genel metrikler, modelin genel başarısı hakkında bilgi verir. Bir sınıflandırma modelinin tüm sınıflar üzerindeki performansını bütüncül olarak değerlendirmeyi amaçlayan ölçütlerdir ve özellikle sınıf dağılımının dengesiz olduğu veri kümelerinde modelin gerçek davranışını anlamak için kritik öneme sahiptir. Genel metriklerde öne çıkan yaklaşımlar: Mikro, makro ve ağırlıklı ortalama gibi hesaplamalardır.

3.5.2.1 Mikro ortalama

Denklem (3.6), (3.7) ve (3.8) üzerinden mikro ortalama hesaplamalarının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Mikro Kesinlik} = \sum_{i=1}^K \frac{TP_i}{TP_i + FP_i} \quad (3.6)$$

$$\text{Mikro Duyarlılık} = \sum_{i=1}^K \frac{TP_i}{TP_i + FN_i} \quad (3.7)$$

$$\text{Mikro F1} = \sum_{i=1}^K \frac{2 \cdot \text{Mikro Kesinlik}_i \cdot \text{Mikro Duyarlılık}_i}{\text{Mikro Kesinlik}_i + \text{Mikro Duyarlılık}_i} \quad (3.8)$$

Mikro-Ortalama metrikler, öncelikle tüm sınıflardaki her sınıftaki doğru pozitif (TP), yanlış pozitif (FP) ve yanlış negatif (FN) sayılarının toplanmasına dayanır. Başka bir deyişle, modelin toplam tahminleri üzerinden hesaplandığı için, veri setindeki örnek sayısının fazla olduğu (çoğunluk) sınıfların performansına daha fazla ağırlık verir (Jin ve diğ., 2012). Bu metrikler, toplam doğru sınıflandırma oranını yansıtmaya yatkın olduğundan, özellikle dengesiz veri setlerinde yüksek sonuçlar verebilirken, azınlık sınıfların kötü performansını maskeleyebilir.

3.5.2.2 Makro ortalama

Denklem (3.9), (3.10) ve (3.11) üzerinden makro ortalama hesaplamalarının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Makro Kesinlik} = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FP_i} \quad (3.9)$$

$$\text{Makro Duyarlilik} = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FN_i} \quad (3.10)$$

$$\text{Makro F1} = \frac{1}{K} \sum_{i=1}^K \frac{2 \cdot \text{Makro Kesinlik}_i \cdot \text{Makro Duyarlilik}_i}{\text{Makro Kesinlik}_i + \text{Makro Duyarlilik}_i} \quad (3.11)$$

Makro-Ortalama metrikler, her bir sınıf için Kesinlik, Duyarlilik ve F1 Skorunun ayrı ayrı hesaplanması ve ardından bu sonuçların aritmetik ortalamasının alınmasıyla elde edilir. Bu yaklaşım, Mikro-Ortalamanın aksine, her bir sınıfa veya yöne eşit ağırlık atar (Jin ve diğ., 2012).

Makro-Ortalama, sınıflardaki örnek sayılarından (frekanstan) bağımsız olduğu için, azınlık sınıfların performansını doğrudan yansıtır ve sınıf dengesizliğinin olduğu durumlarda modelin genel etkinliği hakkında daha net bir resim sunar (Jin ve diğ., 2012). Örneğin, bir modelin nadir görülen bir sınıfı doğru tahmin etme yeteneği, Makro Duyarlilik değeri üzerinde önemli bir etkiye sahip olacaktır.

3.5.2.3 Ağırlıklı ortalama

Denklem (3.12), (3.13) ve (3.14) üzerinden ağırlıklı ortalama hesaplamalarının nasıl gerçekleştirildiği gösterilmektedir.

$$\text{Ağırlıklı Kesinlik} = \sum_{i=1}^K \left(\frac{n_i}{N} \cdot \frac{TP_i}{TP_i + FP_i} \right) \quad (3.12)$$

$$\text{Ağırlıklı Duyarlılık} = \sum_{i=1}^K \left(\frac{n_i}{N} \cdot \frac{TP_i}{TP_i + FN_i} \right) \quad (3.13)$$

$$\text{Ağırlıklı F1} = \sum_{i=1}^K \left(\frac{n_i}{N} \cdot \frac{2 \cdot \text{Ağırlıklı Kesinlik}_i \cdot \text{Ağırlıklı Duyarlılık}_i}{\text{Ağırlıklı Kesinlik}_i + \text{Ağırlıklı Duyarlılık}_i} \right) \quad (3.14)$$

Ağırlıklı Ortalama (Weighted Averaging), Makro ve Mikro ortalamanın sınırlılıklarını dengelemek amacıyla kullanılır. Bu yaklaşım, her bir sınıfın Kesinlik, Duyarlılık ve F1 Skoru metriklerini hesaplar ve ardından bu değerleri, o sınıfa ait örnek sayısının genel veri setindeki oranına göre (yani frekansına göre) ağırlıklandırarak ortalamasını alır. Ağırlıklı F1 Skoru, dengesiz veri setlerinde modelin performansını daha kapsamlı bir şekilde değerlendirmek için sıklıkla Makro ve Mikro F1 skorlarıyla birlikte raporlanır (Jin ve diğ., 2012).



4. UYGULAMA

Tez çalışması kapsamında uygulamalı süreç aşağıdaki temel adımlar üzerinden yürütülmüştür.

- Veri Setinin Hazırlanması
 - Veri Madenciliği
 - Etiketleme
 - Ön İşleme
- Varsayılan Parametreler ile Model Eğitimleri
- Optuna Hiperparametre Optimizasyonu ile Model Eğitimleri
- Sonuçların Değerlendirilmesi
- Modellerin Yayınlanması

4.1 Veri Madenciliği Süreci

Python temelli bir ortam ile veri madenciliği süreci yürütülmüştür. Tüm kütüphaneler ve bağımlılıklar, ayrı bir sanal ortam (venv) üzerinde kurulmuş ve sürümleri sabitlenmiştir. Bu sayede çalışma farklı bilgisayarlarda da aynı koşullar altında tekrarlanabilir hale getirilmiştir. Kodlar Jupyter Notebook ortamında geliştirilmiş olup temel olarak Selenium ve BeautifulSoup kütüphanelerinden yararlanılmıştır. Javascript altyapılı bir e-ticaret platformu üzerinden veri çekim süreci yürütüldüğü için çeşitli kısıtlamaları aşabilmek üzere, Selenium ile otomatik tarayıcı açılarak tıklama, sayfada gezinme, bekleme, imleç hareketleri gibi çeşitli insansı davranışlar modellenmiş ve proxy ip yönetimi organize edilmiştir. İlgili sayfalardan ilgili veriler HTML formatında indirilmiştir, sayfanın tamamı indirilmemiştir. Böylece gereksiz veri ve trafik yükü önlenmiştir.

Yalnızca kamuya açık sayfalar ziyaret edilerek süreç beş aşamadan işletilmiştir: (i) ürün listeleme – filtre alanı, (ii) marka bazlı ürün listeleme, (iii) ürün detay, (iv) ürün

yorumları ve (v) ürün soru–cevapları. İlgili bağlantı linklerine (url) erişmek için kapsamlı bir şekilde 4 farklı site sayfa yapısı üzerinden sırayla sistematik olarak veri çekilmiştir.

(i) ürün listeleme – filtre alanı: İlk aşama olarak “Güneş Gözlüğü” üst kategorisi için “Yüksek Puanlı Ürünler” ve “Yüksek Puanlı Satıcılar” filtreleri uygulanmıştır. İlgili site, belirli bir noktadan sonra yeni ürünleri yüklemeyi durdurduğu için, yeterli ölçüde veri seti oluşturabilmek üzere filtre paneli dışa aktarılmış ve buradan marka-özel ürün listeleme bağlantı linkleri elde edilmiştir.

(ii) marka bazlı ürün listeleme: Her markanın ürün listeleme sayfası açılarak ürün kartları elde edilmiştir. Her ürün kartında yer alan ürün detay sayfa bağlantı linkleri belirlenmiş ve ürünün adı, fiyatı, puanı gibi temel bilgiler bu aşamada toplanmıştır.

(iii) ürün detay: Bu aşamada her marka içerisindeki her ürün için ilgili ürün detay sayfasında ürün adı, fiyatı, puanı, özellikleri, açıklamaları ve benzer ürünleri gibi daha detaylı ürün bilgileri toplanmıştır. Sonraki aşamalarda kullanılmak üzere ürün yorumlar ve soru-cevap sayfa bağlantı linkleri elde edilmiştir.

(iv) ürün yorumları: İlgili ürün için yorum metni, yorum puanı ve yorum tarihi gibi yorumlarla ilgili veriler toplanmıştır.

(v) ürün soru–cevapları: İlgili ürün ve satıcısı için tarih, soru metni ve cevap metni gibi veriler toplanmıştır.

Tüm bu işlemler sırasında, siteye gereksiz istek göndermemek ve sistemi yormamak için çeşitli optimizasyonlar uygulanmıştır. Sayfa kaydırmaları sınırlı tutulmuş, her kaydırmadan sonra kısa beklemler eklenmiştir. Yeni içerik yüklenmediği durumda süreç otomatik olarak durdurulmuş ve başarısız yüklemelerde sınırlı sayıda yeniden deneme yapılmıştır. Her sayfa için Xpath seçiciler ile yalnızca gerekli bölümün HTML çıktısı alınmış, sayfanın tamamı indirilmemiştir. Böylece gereksiz veri ve trafik yükü önlenmiştir. Elde edilen HTML verileri, Beautiful Soup kullanılarak CSS seçiciler yardımıyla işlenmiş, anlamlı sütunlara ayrılmış ve CSV dosyaları halinde kaydedilmiştir. Böylece, dinamik web sayfalarındaki veriler kontrollü, sistematik ve tekrarlanabilir bir şekilde elde edilmiştir.

Toplanan veriler üzerinde uygun formata dönüştürmek üzere çeşitli veri manipülasyon işlemleri uygulanmıştır. Böylelikle rakam, ondalık sayı, metin ve url gibi veri tiplerine uygun olarak veri setleri oluşturulmuştur.

Veri toplama süreci, belirlenen kapsam ve araştırma hedefleri doğrultusunda yürütülmüştür. İlgili e-ticaret platformu üzerindeki “Güneş Gözlüğü” kategorisinde yer alan markalar, ürünler ve kullanıcı etkileşimleri (yorumlar ve soru-cevaplar) belirli bir çeşitlilik ve temsil düzeyine ulaşana kadar tarama işlemleri sürdürülmüştür. Bu tarama süreci birkaç hafta içerisinde yavaş ve insansı bir davranış ile yürütülmüş olup ilgili e-ticaret platformunun sunucularına fazla yük oluşturulmamıştır. Veri hacmi, model eğitimi ve analiz aşamalarında istatistiksel olarak anlamlı sonuç üretebilecek seviyeye ulaştığında, sürece son verilmiştir. Böylece hem kaynak sisteme gereksiz yük oluşturulmamış hem de veri madenciliği faaliyetleri, çalışmanın amaçlarına uygun ve yeterli bir noktada dengelenmiştir.

4.2 Veri Setlerine Genel Bakış

Veri toplama süreci ile 7 farklı tablo veri seti oluşturulmuş olup alt başlıklar üzerinden detaylandırılmaktadır. Bu kapsamda veri setleri; ürün detay bilgileri, ürün yorumları ile ürünlere ilişkin soru ve cevap içeriklerini kapsayacak şekilde yapılandırılmıştır. Söz konusu alt başlıklar, farklı URL yapıları ve web sayfası şablonları üzerinden veri madenciliği teknikleri kullanılarak elde edilmiştir. Ürün detay verileri daha çok statik ve tanımlayıcı bilgiler içerirken, ürün yorumları ve soru-cevap bölümleri kullanıcı deneyimine dayalı dinamik metinler sunmaktadır. Tez çalışması kapsamında ise duygu analizi görevine doğrudan katkı sağlaması nedeniyle yalnızca ürün yorumlarına ait metin verileri kullanılmış, diğer veri setleri bağlamsal bütünlüğü destekleyici nitelikte değerlendirilmiştir. Bu kapsamda Çizelge 4.1’de veri setilerindeki başlıca önemli özet bilgiler ve çeşitlilik ifade edilmektedir.

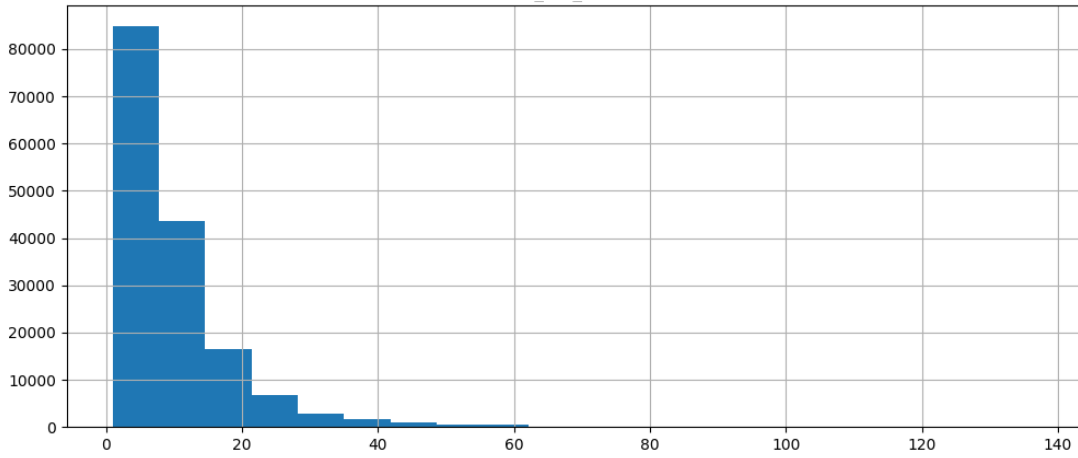
Çizelge 4.1 : Veri setleri analizi

Veri	Sayısı
Ürün	11.014
Marka	242
Satıcı	625
Yorum	159.552
Soru & Cevap	55.551

Çizelge 4.1’de görüldüğü üzere, veri setinin 11.014 farklı ürün, 242 farklı marka ve 625 farklı satıcıyı kapsaması, yorum metinlerinin yalnızca nicel değil, aynı zamanda anlamsal ve bağlamsal açıdan da yüksek düzeyde çeşitlilik içermesini sağlamıştır. Farklı ürün kategorileri, marka konumlandırmaları ve satıcı profilleri üzerinden

üretilen kullanıcı geri bildirimleri, duygu ifadelerinin kullanım biçimlerinde belirgin yapısal farklılıkların ortaya çıkmasına katkı sunmuştur. Bu çeşitlilik sayesinde model, farklı bağlamlarda ifade edilen pozitif, negatif ve nötr duygu örüntülerini daha etkin biçimde öğrenebilmekte ve genellenebilir bir temsil yeteneği kazanmaktadır.

Şekil 4.1’de görüldüğü üzere yorumlardaki kelime sayısı geniş bir aralıkta değişkenlik göstermekle birlikte, dağılımın üstel bir yapı sergilediği gözlemlenmektedir. Kullanıcı yorumlarının büyük bir bölümü kısa ve öz ifadelerden oluşurken, daha sınırlı sayıda yorumun ise ayrıntılı ve uzun açıklamalar içerdiği görülmektedir.









Şekil 4.1 : Yorumlardaki kelime sayıları dağılımı

Model eğitimleri sürecinde, özellikle derin öğrenme ve transformer tabanlı modellerin sabit girdi uzunluğu kısıtları dikkate alındığında, bu tür bir dağılımın doğrudan modelleme stratejilerini etkilediği görülmektedir. Bu nedenle, yorum metinlerinin uygun şekilde kırpılması, normalize edilmesi veya belirli eşik değerler doğrultusunda ele alınması, hem hesaplama verimliliği hem de model performansı açısından önem arz etmektedir. Bu kapsamda elde edilen veriler, sonraki aşamalarda çeşitli ön işleme ve eleme süreçlerine tabi tutularak, model eğitime uygun ve tutarlı bir girdi yapısı oluşturulması hedeflenmiştir.

4.2.1 Ürün detay

Ürün id, marka, sku, adı, fiyatı, açıklaması, değerlendirme ortalaması, değerlendirme sayısı, yorum sayısı, soru ve cevap sayısı, ürün özellikleri, benzer ürün id'leri, sepete ekleme sayısı, favorileme sayısı, satış sayısı, ürün seo başlığı, ürün seo açıklaması, görsel linkleri, diğer linkler vb gibi ürüne dair veri bilgileri elde edilmiştir. Şekil 4.2’de ilgili veri seti için örnek bir gösterim sunulmaktadır.



Görsel	Ürün Adı	Marka	Fiyat	Değerlendirme Puanı	Değerlendirme Sayısı	Yorum Sayısı	Renk	Çerçeve Formu	Çerçeve Tipi	Çerçeve Materyali	Çerçeve Renk	Cam Tipi	Cam Materyali	Cam Renk	Menşei	Desen	Ekartman	Özellik	Ek Özellik
	Benx Sunglasses; Benx 9059; C03 52-19 145 Erkek Güneş Gözlüğü	Benx Sunglasses	₺1.399	5,00	2	2	Çok Renkli	Dikdörtgen	Çerçevesiz	Kombine	Çok Renkli	Degrade	Uv400	Füme		Düz		Dikdörtgen	
	Benx Sunglasses; Benx Unisex; Güneş Gözlüğü 9055 C01	Benx Sunglasses	₺1.800	5,00	1		Gri	Dikdörtgen	Çerçevesiz	Kemik	Siyah	Degrade	Uv400	Füme	Tr	Degrade		Dikdörtgen	
	Benx Sunglasses; Benx 9027 -; D343 Polarize Kadın Güneş Gözlüğü	Benx Sunglasses	₺1.850	5,00	1	1	Siyah	Yuvarlak	Çerçevesiz	Kemik	Siyah	Degrade	Polarize	Çok Renkli		Düz			
	Benx Sunglasses; Benx 9027; Dm05 Unisex Güneş Gözlüğü	Benx Sunglasses	₺1.750	4,50	2	1	Siyah	Oval	Çerçevesiz	Plastik	Altın	Mat	Uv400	Kahverengi		Düz		Uzun	
	Benx Sunglasses; Benx 9054; Unisex Kare Güneş Gözlüğü	Benx Sunglasses	₺1.850	5,00	1		Siyah	Geometrik	Çerçevesiz	Plastik	Siyah	Degrade	Uv400	Siyah	Tr	Düz	65	Uzun	Yıkamalı
	Benx Sunglasses; Benx 9059 C06; 52 Güneş Gözlüğü	Benx Sunglasses	₺1.410	5,00	2	2	Kahverengi	Dikdörtgen	Çerçevesiz	Kombine	Şeffaf	Degrade	Polarize	Kahverengi	Tr	Degrade		Dikdörtgen	

Şekil 4.2 : Ürün detay

4.2.2 Ürün yorumları

Ürün id, yorum tarihi, yorum yapan kişi, yorum yapan kişi tipi, yorum yapan kişi kilo, yorum yapan kişi boy, satıcı, puan, yorum, yorum beğeni sayısı, yorum görselleri vb gibi yorumlara dair veri bilgileri elde edilmiştir. Çizelge 4.2’de orjinal haliyle (ön işlemez) yorumlardan örnekler gösterilmektedir.

Çizelge 4.2 : Ürün yorumları

Puan	Yorum
5	çok memnun kaldım yüze tam oturuyor rahatsız etmiyor tavsiye ederim
5	yorumları inceleyerek sipariş verdim çok memnun kaldım.. kullanışlı hafif kaliteli ve çok tarz..
4	yazları gözümde çıkırmıyrm baya bı havalı duruyo biraz büyük ama olsun
4	gözlüğü araç sürerken kullanmak için aldım yolda denemedim ama güneşte takip baktım berrak bir görüntüsü var çerçevesi mat olsaydı biraz daha şık duracaktı belki
3	Çok büyük benim yüzüme gitmedi
3	Camları adi duruyor veya ben kendime yakıştıramadım bilmiyorum fiyatına göre baya iyi ama
2	yüzüme çok büyük olduğu için iade
2	iki adet gelecekti tek adet geldi
1	Kalitesini beğendim diyemem
1	Ürün bana çok kaliteli gelmedi açıp kaparken plastik sesi geliyo mendilde çıkmadı içinden

4.2.3 Ürün soru ve cevapları

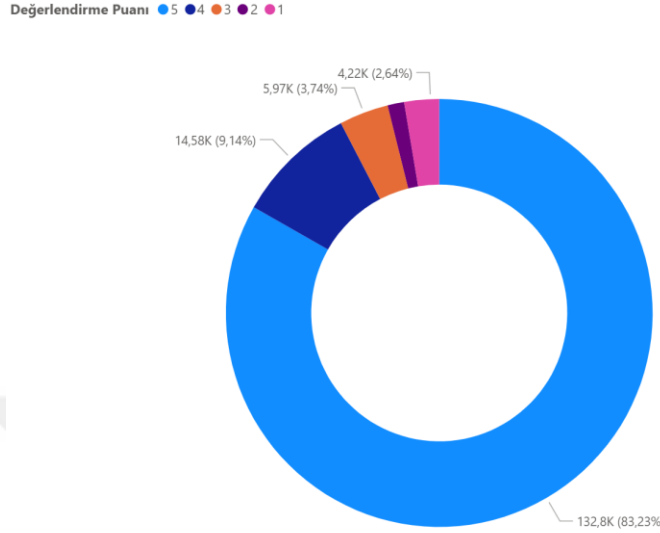
Ürün id, soru cevap tarihi, soru soran kişi adı, ürünün satılma durumu, soru metni, cevap metni, beğeni sayısı, satıcı ünvanı, cevaplanma süresi vb gibi soru ve cevap ile ilgili bilgiler elde edilmiştir. Çizelge 4.3’de orjinal haliyle soru ve cevaplardan örnekler gösterilmektedir.

Çizelge 4.3 : Ürün soruları ve cevapları

Soru	Cevap
Cam rengi kahverengi mi altın rengi mi	Merhaba cam rengi yeşil kahvedir
Gözlüğü almak istiyorum fakat çerçeve rengi sedef yazıyor ama fotoğraflı yorumlarda siyah sedef çerçeve hangi renk oluyor	Merhaba, ürün görseldeki gibi gönderilmektedir. İlginiz için teşekkür ederiz.
kaç ekartman?	Merhaba, Ekartman: 56'dır. Keyifli alışverişler dileriz.
bana olurmu	Bilmiyoruz ki olur mu?
Merhaba ürün orjinal mi evet veya hayir demeniz yeterli olacaktır çünkü daha önce bu soruya verdiğiniz cevap beni tatmin etmedi şimdiden teşekkürler	Merhaba, platformumuzda satışa sunulan ürünlerin orijinallik kontrolü, ürünler satışa çıkmadan önce yapılıyor. Alışverişinizden sonra aldığınız ürünle ilgili bir sorunuz olursa bizimle hemen iletişime geçebilirsiniz. Gerekli kontrolleri mümkün olan en kısa sürede yapıp size yardımcı olmaktan mutluluk duyarız. Keyifli alışverişler dileriz.
hemen tükeniyor hemen stok açsanızda alsak	Merhaba sınırlı sayıda geliyor talep fazla olduğu için hemen tükenmektedir . Takipte kalmanızı rica eder ilginiz için teşekkür ederiz :)

4.3 Veri Etiketleme

Veri madenciliği sonrasında Şekil 4.3’de görüleceği üzere 160 bin yorum ile dengesiz bir veri seti elde edildi.



Şekil 4.3 : Tüm yorumların puan dağılımı

Negatif ve nötr yorumların pozitif yorumlara göre görece daha az olması e-ticaret platformlarında beklenen bir durum olarak kabul edilmektedir. Etiketleme sürecinde bu dengesizliği minimize etmek amacıyla nötr ve negatif yorumların büyük çoğunluğu, pozitif yorumların ise bir kısmı titizlikle etiketlenmiştir (Çizelge 4.4).

Çizelge 4.4 : Yorum, puan ve etiket dağılımları

Puan	Yorum Sayısı	Etiketlenen Yorum Sayısı
5	132.803	43.402
4	14.577	5.068
3	5.966	5.826
2	1.990	1.938
1	4.216	4.063
Toplam	159.552	60.297

İnsan hatası olabileceği düşüncesiyle etiketlenen yorumların kontrolü amacıyla jupyter notebook ortamında Ollama büyük dil modellerden Gemma3:4b ve LLaMa3.2:3b open ai api entegrasyonu kullanılarak tahminleme yapılmıştır.

Tahminlemelerde skora yapararak 1101 etikette revizyon gerekleřtirilmiřtir. Yorum puanı karřılıđı, manuel etiket, Gemma3:4b tahmini ve llama3.2:3b tahmini ađırlıklı puanlama ile nihai etiketleme gerekleřtirilmiřtir.

Denklem (4.1) ve (4.2) zerinden tahminleme skora hesaplamalarının nasıl gerekleřtirildiđi gsterilmektedir.

$$C = \{\text{pozitif, ntr, negatif}\} \quad (4.1)$$

Drt ayrı kaynaktan gelen etiketler:

- h: manuel etiketi
- g: Gemma etiketi
- l: LLaMA etiketi
- r: puan etiketi

Her sınıf iin skor:

$$S(c) = 2.1(h = c) + 1.1(g = c) + 1.1(l = c) + 1.1(r = c), c \in C \quad (4.2)$$

řeklinde hesaplanmaktadır. Sonrasında maksimum skorlu sonu nihai etiket olarak yazılmaktadır. Eđer eřitlik olması durumunda ise mevcut etiket korunmaktadır.

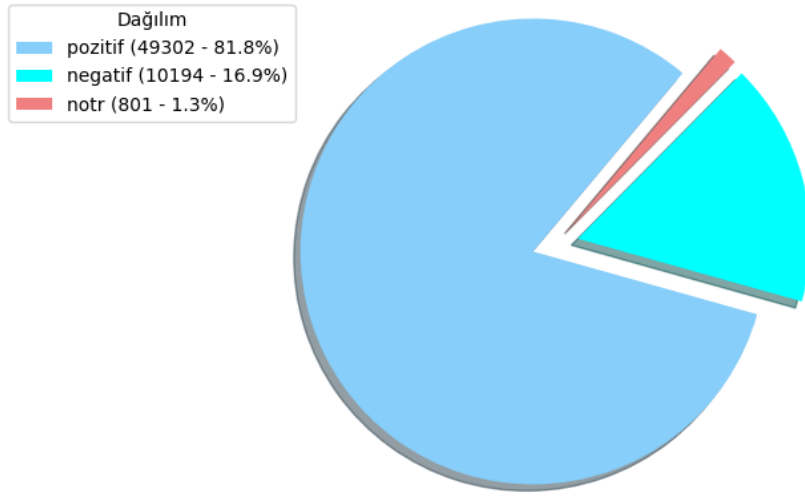
Bu revizyon iřlemi izelge 4.5 zerinden rneklerle gsterilmekte olup, n iřeme adımlarına tabi tutulan yorumlar zerinden model tahminlemesi yapılarak yrtlmřtir.

rnek olarak, “Gzel bir gzlk fakat yzme byk geldiđi iin iade” yorumunda, metin ierisinde hem olumlu hem de olumsuz duygu ifadeleri birlikte yer almaktadır. “Gzel bir gzlk” ifadesi rne ynelik pozitif bir deđerlendirme sunarken, “iade” vurgusu kullanıcı memnuniyetsizliđine iřaret eden gl bir olumsuz sinyal oluřturmaktadır. Ayrıca 3 puanlı deđerlendirme, puan etiketi aısından ntr bir karřılık retmiřtir. Bu nedenle manuel etiket pozitif olarak belirlenmiř olsa da, Gemma3:4b ve LLaMA3.2:3b modelleri yorumu negatif olarak sınıflandırmıřtır. Denklem (4.2) ile hesaplanan ađırlıklı skorlar sonucunda negatif sınıf en yksek deđer almıř ve incelemeye tabi tutulması gerekliliđini vurgulamıřtır. Bu kapsamda tekrar manuel olarak etiketlenerek nihai etiket negatif olarak gncellenmiřtir.

Çizelge 4.5 : Yorum etiket düzenlemeleri

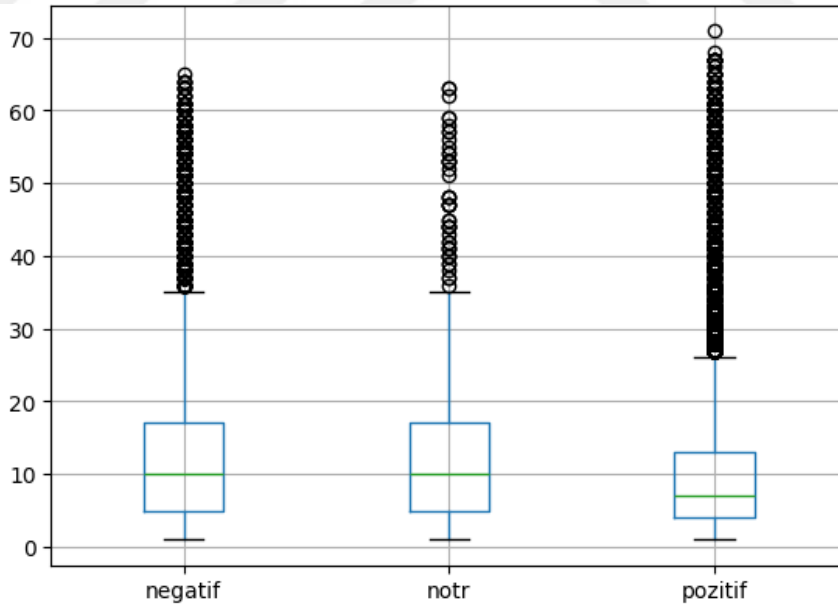
Puan	Yorum	Duygu Etiketi	Gemma3:4b Tahmini	Llama3.2:3b Tahmini	Puan Etiketi	Pozitif Skor	Negatif Skor	Nötr Skor	Nihai Duygu Etiketi
3	Güzel bir gözlük fakat yüzüme büyük geldiği için iade	pozitif	negatif	negatif	nötr	2	2	1	negatif
4	Orijinal gibi ama emin değilim tabii ki gayet şık duruyor beğendim .	nötr	pozitif	pozitif	pozitif	3	0	2	pozitif
3	Orjinal olduğundan emin değilim	negatif	nötr	nötr	nötr	0	2	3	nötr
5	Güzel gözlük sadece biraz ağır kullanımı çok rahat değil	nötr	negatif	negatif	pozitif	1	2	2	negatif
4	Yamuk geldi gidip optikte düzelttirdim onun dışında hiçbir sorunu yoktu	negatif	pozitif	pozitif	pozitif	3	2	0	pozitif
1	Farklı bir satıcı da 1500 lira daha uygun	nötr	negatif	negatif	negatif	0	3	2	negatif

Nihai etiketleme sonrasındaki yorumların duygu etiket dağılımları Şekil 4.4'de görüldüğü gibidir.



Şekil 4.4 : Etiketlenen yorumların duygu etiketi dağılımı

Kısmen dengesiz bir veri seti olarak dağılım göstermektedir. Kutu Grafik Analizi ile üç farklı duygu sınıfındaki (negatif – nötr – pozitif) yorumların kelime sayısı dağılımları Şekil 4.5'de görüldüğü üzere karşılaştırılmaktadır.



Şekil 4.5 : Etiketlenen yorumların kutu grafik analizi

Bu grafik yorumlandığında:

- Yorum yapan kişiler, nötr ya da negatif yorum yazarken ortalama olarak benzer uzunluklarda yazmaktadırlar.

- Pozitif yorumların biraz daha kısa olma eğiliminde olduğu gözlemlenmektedir.
- Aykırı uzunlukta yorumlar bulunmakla birlikte, görece genele göre daha azdır. Aykırı uzunluktaki yorumlarda memnun kullanıcılar daha uzun yorum yazma eğilimindedir.

Ön işleme ve filtrelemeler gibi sonraki adımların uygulanması ile veri seti nihai halini alarak duygu etiketi dağılımı kısmen iyileşmektedir.

4.4 Ön İşleme

Model eğitimine uygun temiz ve tutarlı bir metin yapısı elde edebilmek için her bir kullanıcı yorumuna sıralı bir ön-işleme hattı (preprocessing pipeline) uygulanmıştır. Böylelikle model eğitimi bağlamsal olarak daha tutarlı ve sağlıklı olacaktır. Bu kapsamda uygulanan çeşitli ön işleme adımları listelenmektedir:

- Normalizasyon
- Küçük harflere çevirme
- URL, etiket (hashtag), email gibi gürültüleri temizleme
- Türkçe karakter düzenlemeleri
- Emoji dönüştürme
- Harf tekrarlarını azaltma
- Noktalama, işaret ve boşluk düzenlemeleri

Düzenlemeler sonrasında ise model eğitimlerinde kullanılacak veri setinde çeşitli filtrelemeler gerçekleştirilmiştir:

- Türkçe dışı yorumlar,
- Tekrar eden yorumlar,
- Konu dışı yorumlar

Böylelikle de benzersiz, zengin ve dil modelleri için daha doğru yapıda bir veri seti elde edilmiştir. Duygu etiketi modeli eğitimleri sürecinde bağlam daha anlamlı ve kararlı olmuştur.

4.4.1 Normalizasyon

Normalizasyon işlemi için çeşitli kütüphaneler ve metodolojiler araştırılmıştır:

- ÚFAL at MultiLexNorm 2021: Kullanıcı tarafından üretilen gürültülü metinlerin otomatik olarak normalleştirilmesine yönelik geliştirilen bir yöntemdir. Bu yaklaşım, sosyal medya gönderileri, kullanıcı yorumları ve çok dilli kısa metinlerde sık karşılaşılan yazım hataları, harf tekrarı, kısaltmalar ve noktalama tutarsızlıklarını düzelteren seq2seq tabanlı bir karakter düzeyi normalizasyon modeli (ByT5) içermektedir. Yöntem, MultiLexNorm 2021 yarışma kapsamında yüksek performans elde etmiş ve özellikle Türkçe, İngilizce, Almanca, Çekçe gibi dillerde gürültü azaltma başarısı ile öne çıkmıştır.
- VNLP: Türkçe metin işleme süreçlerinde kullanılan bir ön-işleme kütüphanesi olarak doğal dil verisini daha tutarlı ve analiz edilebilir bir forma dönüştürmeyi amaçlar.
- Zemberek: Türkçe doğal dil işleme alanında en bilinen ve en uzun süreli açık kaynak projelerden biridir. İlk sürümü 2007 yılında yayımlanan Zemberek, özellikle morfolojik çözümleme, heceleme, yazım denetimi ve kök-ek ayrıştırma gibi Türkçe'ye özgü dilsel işlemleri desteklemek üzere geliştirilmiştir. Türkçe'nin eklemeli yapısı için özel olarak tasarlanmış geniş kural tabanlı yapısı sayesinde metin ön-işleme aşamalarında güvenilir bir temel sunmaktadır. Yıllar içinde hem topluluk desteği hem de akademik kullanımı artan Zemberek, günümüzde Türkçe DDİ çalışmalarında morfolojik analiz için en sık başvuru alan açık kaynak araçlardan biri olmaya devam etmektedir.

Bu yöntemlerden normalizasyon işlemi için VNLP ve Zemberek kütüphaneleri kullanılarak denemeler yapılmış ve tatmin edici sonuçlara ulaşılamamıştır. Bu kapsamda W-NUT 2021: Multilingual Lexical Normalization (MultiLexNorm 2021) yarışmasında (Beck et al, 2021) birincilik elde eden ÚFAL modeli kullanılmaya karar verilmiştir. Bu model ByT5 tabanlı olup; önce sentetik verilerle ön-eğitime tabi tutulmuş, ardından gerçek normalizasyon verileriyle ince ayar yapılmış bir modeldir (ÚFAL, 2021). Ön eğitim ve ince ayar aşamalarında çok dilli veri seti kullanıldığı, bu veri setleri içerisinde Türkçe diline de yer verildiği ifade edilmiştir (ÚFAL, 2021).

Çalışmada kullanılan normalizasyon fonksiyonu, metinleri küçük harfe dönüştürme, gereksiz karakterleri temizleme, tekrar eden harfleri sadeleştirme, noktalama işaretlerini standartlaştırma ve temel biçim uyumunu sağlama gibi adımlar içererek, model eğitiminde daha kararlı ve gürültüsüz bir veri yapısı elde edilmesine katkı sağlamıştır

4.4.2 Küçük harflere dönüştürme

Yorumlardaki büyük/küçük harf tutarsızlıklarını gidermek amacıyla, yorumlar Türkçe'ye özgü karakter yapısı dikkate alınarak küçük harfe dönüştürülmüştür. Bu işlem sırasında Unicode normalizasyonu uygulanmıştır. Böylece model eğitiminde gereksiz varyasyonların önüne geçilmiştir.

4.4.3 Gürültü azaltma

Modelin öğrenmesini olumsuz etkileyen URL, e-posta adresi ve kullanıcı etiketleri gibi gürültülü öğeler metinden çıkartılmıştır. Bu öğeler tamamen silinmek yerine “url”, “email”, “user” gibi nötr terimlerle değiştirilmiş ve böylece hem bilgi kaybı azaltılmış hem de modelin dikkat dağıtıcı unsurlardan arındırılması sağlanmıştır. Böylelikle model eğitimleri sürecinde duygu bağlamları iyileştirilmiştir.

4.4.4 Türkçe karakter düzenlemeleri

Sosyal medya ve e-ticaret yorumlarında sık görülen ASCII karakter kullanımı nedeniyle bozulan Türkçe kelimeler, güvenli bir “deasciify” fonksiyonu ile düzeltilmiştir. Sadece tamamen ASCII karakterlerden oluşan kelimelere dönüştürme uygulanmış ve Türkçe harf içeren kelimeler korunmuştur. Böylece yanlış düzeltmelerin önüne geçilmiş ve metnin Türkçe morfolojik yapısı iyileştirilmiştir.

4.4.5 Emoji dönüştürme

Yorumlarda sıklıkla kullanılan emojiler, doğrudan kaldırılmak yerine anlamlarını temsil eden metinsel açıklamalara dönüştürülmüştür (örneğin “👉” → “emoji_kalp_gözlü_gülümseme”). Bu işlem için özel bir emoji–anlam sözlüğü kullanılmıştır (Kamali, 2024). Böylece duygusal içeriği taşıyan emojiler, modelin öğrenebileceği biçimde korunmuştur. Duygular korunurken, bağlam iyileştirilmiştir.

4.4.6 Harf tekrarlarını azaltma

“çooook”, “müüüüthiş”, “harikAAAA” gibi kullanıcıların duygu yoğunluğunu ifade etmek için tekrarladığı karakterler, iki tekrar ile sınırlandırılarak normalize edilmiştir. Rakam içeren tekrarlar (ör. ürün fiyatları) hariç tutulmuştur. Bu işlem ile modelin anlamsal olarak aynı kelimeyi farklı varyasyonlarla görmesini engelleyip veri tutarlılığının artırılması amaçlanmıştır.

4.4.7 Noktalama, işaret ve boşluk düzenlemeleri

Yorumlarda yer alan aşırı noktalama işaretleri (“!!!”, “???”) sadeleştirilmiş, para sembolleri tek bir “tl” şeklinde normalize edilmiştir. Ayrıca ardışık boşluklar tek boşluğa dönüştürülmüştür. Böylece model eğitimi sırasında gereksiz sembolik varyasyonların etkisi azaltılmıştır.

4.4.8 Yorum dil tespiti

Yorumların içerisinde Türkçe dışındaki dillerde yazılmış örneklerin bulunması, model eğitim sürecini olumsuz etkileyebileceği için veri kümesinde dil temizliği yapılmıştır. Bu amaçla Facebook AI Research tarafından geliştirilen ve 176 dil üzerinde eğitilmiş olan FastText Language Identification modeli (Facebook AI Research, 2017) kullanılmıştır. FastText’in karakter-n-gram temelli yaklaşımı sayesinde kısa metinlerde dahi yüksek doğrulukla dil tahmini yapılabilmiş ve yorumların büyük bölümünün dili güvenilir bir şekilde belirlenmiştir. Çalışmada her bir yorum için FastText çıktısı ve modelin tahmin olasılığı alınmış, belirlenen eşik değerinin altında kalan veya Türkçe dışındaki dillere ait olduğu tespit edilen yorumlar veri kümesinden çıkartılarak temizlenmiştir. Böylelikle eğitim verisi homojen bir dil yapısına kavuşturulmuş, modelin öğrenme süreci gürültüden arındırılmıştır.

4.5 Model Eğitimleri

Model eğitimleri öncesinde veri setinde çeşitli ön işleme adımlarının gerçekleştirilmiş olup nihai veri seti oluşturulmuştur. Bu süreçte:

- Eğitim, doğrulama ve test alt veri setlerinin oluşturulması
- Temel model araştırmaları
- Varsayılan parametreler ile model eğitimleri

- Optuna Hiperparametre optimizasyonu ile model eğitimleri
- Sonuçların değerlendirilmesi
- Eğitilen modellerin yayınlanması

adımları takip edilmiştir.

Temel model araştırmaları neticesinde başarılı ve güncel modeller seçilmiştir:

- dbmdz/electra-base-turkish-mc4-uncased-discriminator
- dbmdz/bert-base-turkish-128k-uncased
- VRLLab/TurkishBERTweet
- Trendyol/tyroberta
- microsoft/mdeberta-v3-base
- artiwise-ai/modernbert-base-tr-uncased

Varsayılan ve Optuna parametre optimizasyonu ile model eğitimleri tamamlanıp genel ve sınıf bazlı metrikler ile kıyaslamaları gerçekleştirilmiştir. Sonrasında eğitilen modeller, Hugging Face platformunda detaylı açıklamalar ve değerlendirme sonuçları ile birlikte yayınlanmıştır:

- tyroberta-turkish-sentiment-optuna-hpo (msamilim, 2025a)
- mdeberta-v3-turkish-sentiment-optuna-hpo (msamilim, 2025b)
- bert-128k-turkish-sentiment-optuna-hpo (msamilim, 2025c)
- turkishbertweet-turkish-sentiment-optuna-hpo (msamilim, 2025d)
- modernbert-turkish-sentiment-optuna-hpo (msamilim, 2025e)
- electra-turkish-sentiment-optuna-hpo (msamilim, 2025f)

Bu eğitim sürecinde, öncelikli olarak google colab ortamında lisans ücreti ödenerek geliştirmeler yapılmıştır. Tez çalışması kapsamında maliyet ve zaman yönetimi açısından Yerel bir geliştirme ortamı üzerinde yapılandırılmasına karar verilmiştir. Çalışmalar Windows 11 işletim sistemi altında WSL (Windows Subsystem for Linux) kullanılarak Ubuntu tabanlı bir Linux işletim sistemi üzerinden yürütülmüştür. Geliştirme ortamı Conda ile yönetilen izole sanal ortamlar üzerinde kurulmuş, böylece paket bağımlılıkları kontrollü ve tekrarlanabilir şekilde yönetilmiştir. GPU

hızlandırmalı eğitimlerin sağlıklı gerçekleştirilebilmesi için NVIDIA sürücülerini, CUDA desteği ve PyTorch (torch==2.8.0+cu128) sürümünü uyumlu şekilde yapılandırılmıştır. Model geliştirme sürecinde Visual Studio Code kullanılmış, WSL üzerinden çalışan Jupyter eklentisi aracılığıyla not defterleri sorunsuz şekilde işletilmiştir. Kullanılan temel kütüphaneler:

- transformers==4.57.0
- torch==2.8.0+cu128
- datasets==4.2.0
- accelerate==1.10.1
- evaluate==0.4.6
- python==3.11.13

yer almıştır. Tüm ince ayar işlemleri, 8 GB VRAM kapasitesine sahip NVIDIA GeForce RTX 3070 Ti ekran kartı üzerinde GPU hızlandırmalı olarak gerçekleştirilmiş; bu donanım yapılandırması hem eğitim sürelerini önemli ölçüde kısaltmış hem de yüksek boyutlu modellerin verimli şekilde optimize edilmesine imkân tanımıştır. Bu bütünleşik yapı sayesinde, yerel bir araştırma ortamında yüksek performanslı ve tekrarlanabilir bir deney altyapısı oluşturulmuştur.

4.5.1 Nihai veri seti

Öncelikle veri madenciliği yöntemleri kullanılarak 159.552 adet ürün yorumu elde edilmiş, ardından etiketleme aşamasında sınıflar arasındaki belirgin dengesizliği azaltmak amacıyla nötr ve negatif yorumların büyük bir bölümü ile pozitif yorumların belirli bir kısmı titizlikle etiketlenerek 60.297 yorumdan oluşan ara bir veri kümesi oluşturulmuştur. Devam eden ön işleme ve filtreleme adımları kapsamında tekrar eden, Türkçe olmayan ve konu dışı içerikler veri setinden çıkarılmış; böylece model eğitimlerinde kullanılmak üzere 50.160 yorumdan oluşan nihai veri seti elde edilmiştir. Şekil 4.6' da sunulan duygu etiketi dağılımı, uygulanan bu adımlar sonucunda sınıf dengesizliğinin önceki duruma kıyasla kısmen azaltıldığını ve veri setinin model eğitimi için daha uygun bir yapıya kavuştuğunu göstermektedir.

Çizelge 4.6 (devamı) : Nihai yorumlar

Yorum	Ön İşleme Sonrası Yorum
tam istedigim gibi geldii t3sekkurler	tam istediğim gibi geldi teşekkürler
Orjinal değil arkadaşlar iade edecem Neden degidirmiyorsunuz O kadar çakma duruyo ki hemen kırılacak gıbyı resmen parasını gram haketmiyor iade	orjinal değil arkadaşlar iade edeceğim neden değiştirmiyorsunuz o kadar çakma duruyor ki hemen kırılacak gibi resmen parasını gram hak etmiyor iade
Guzel ama küçük okadar abartıldığı gıbyı bısey degil iade	güzel ama küçük o kadar abartıldığı gibi bir şey değil iade
hep istediğim bir ürünü orjinal çok güzel bence herkese yakışacak bi model	hep istediğim bir ürünü orjinal çok güzel bence herkese yakışacak bir model

Veri setindeki gerçek yorumlardan örnekler incelendiğinde, uygulanan ön işleme sürecinin genel olarak başarılı, tutarlı ve model eğitime uygun bir metin yapısı ürettiği gözlemlenmiştir. Özellikle Türkçe’ye duyarlı küçük harfe dönüştürme, harf tekrarlarının azaltılması ve ASCII karakterli kelimelerin Türkçe biçimlerine dönüştürülmesi adımları, “çooooom → çok”, “muhtesem → mühteşem”, “çpk → çok”, “t3sekkurler → teşekkürler” gibi örneklerde oldukça etkili sonuçlar vermiştir. Bunun yanında emoji’lerin anlamlı metin karşılıklarına dönüştürülmesi (“🤩 → emoji_yıldız_çarpmış”, “😊 → emoji_3_kalpli_gülümseyen_yüz”) yorumların duygusal tonunun korunarak modele aktarılmasını sağlamış ve bu yönüyle önemli bir normalizasyon katkısı sağlamıştır. Ayrıca noktalama ve boşluk düzenlemeleri sonrasında tüm cümlelerin daha okunabilir, standart ve analiz edilebilir bir hâle geldiği görülmektedir.

Bununla birlikte, bazı örneklerde normalizasyonun tam başarı sağlayamadığı sınırlı durumlar göze çarpmaktadır. Örneğin “bi model → bir model” dönüşümü doğru yapılmış olsa da, “bir ürünü → bir ürünü” gibi beklenmedik bir dönüşüm gerçekleşmiştir. Benzer şekilde yazım hatası çok yüksek olan “orjinal → orijinal” dönüşümü çoğunlukla doğru yapılmış olsa da, kimi kelimelerde kök tespitinin zayıf kaldığı görülmektedir.

Genel olarak değerlendirildiğinde, ön işleme süreci büyük çoğunlukla doğru, güvenilir ve modellerin performansını artıracak şekilde metni sadeleştirdiği; kalan küçük hataların ise gerçek dünya verisinin doğasından kaynaklanan kabul edilebilir bir

seviyede olduğu görülmektedir. Bu nedenle uygulanan ön işleme yaklaşımı, çalışma kapsamında hedeflenen duygu analizi modelleri için sağlam bir temel oluşturmuştur.

Pozitif sınıfa ait yorumlar için oluşturulan kelime bulutu Şekil 4.7’de gösterilmektedir.



Şekil 4.7 : Pozitif yorumların kelime yoğunluk bulutu

Özellikle “güzel”, “beğendim”, “harika”, “kaliteli”, “şık”, “mükemmel”, “teşekkür ederim” ve “tavsiye ederim” gibi olumlu nitelik bildirci kelimelerin yoğunlukla yer aldığı görülmektedir. Bu durum, pozitif yorumların çoğunlukla ürünün estetik görünümü, kalite algısı ve kullanım rahatlığı üzerinde yoğunlaştığını göstermektedir. Ayrıca “gözlük”, “ürün” ve “model“ gibi temel kategori ifadelerinin sıklıkla tekrarlandığı; bunun yanında “sıcıcı”, “kargo”, “paketleme” ve “elime ulaştı” gibi alışveriş sürecine ilişkin ifadelerinin de pozitif geri bildirimlerde önemli bir yer tuttuğu gözlemlenmiştir. Bu dağılım, kullanıcıların yalnızca ürün özelinde değil, aynı zamanda teslimat hızından ve satıcı güvenilirliğinden de olumlu şekilde etkilendiklerini ortaya koymaktadır.

Negatif sınıfa ait yorumlar için oluşturulan kelime bulutu Şekil 4.8’de gösterilmektedir.



Şekil 4.9 : Nötr yorumların kelime yoğunluk bulutu

Yorum yapan kişilerin ne tamamen olumlu ne de tamamen olumsuz bir değerlendirme içeren yorumlarında hangi temalara odaklandığını göstermektedir. Görselde “gözlük”, “ürün”, “biraz”, “küçük”, “fiyat”, “iyi”, “uygun”, “siyah”, “idare eder”, “göre” ve “renge” gibi kelimelerin daha ön planda olduğu görülmektedir. çok açıklayıcı, tanımlayıcı ve nesnel ifadelerin ön plana çıktığı görülmektedir. “Fakat”, “ancak”, “küçük” ve “büyük” gibi karşılaştırmalı ifadeler ise nötr yorumlarda sıkça görülen hafif olumlu-hafif olumsuz değerlendirmelerin bir arada kullanıldığını göstermektedir. Genel olarak, bu kelime dağılımı nötr yorumların ürün özelliklerine ilişkin daha açıklayıcı ve bilgi aktarımına dönük bir yapıda olduğunu göstermektedir.

4.5.2 Eğitim, doğrulama ve test ayrımı

Duygu analizi modellerinin eğitimi öncesinde, nihai veri seti modelleme sürecine uygun hale getirilmek amacıyla yeniden yapılandırılmış; tablo sütun isimleri “metin-etiket” biçimine indirgenerek yalnızca ön işleme adımlarından geçirilmiş yorum metinleri ile bunlara karşılık gelen üç sınıflı duygu etiketlerinin sayısal temsilleri (negatif: 0, nötr: 1, pozitif: 2) veri setinde tutulmuştur. Bu sadeleştirme işlemi, model girişlerinin tutarlı ve standart bir formatta sunulmasını sağlayarak eğitim sürecinin daha verimli ve tekrarlanabilir biçimde yürütülmesine katkı sağlamıştır.

Eđitim s¼recinde kullanılmak üzere hazırlanan nihai veri seti dađılımları Çizelge 4.7’de gösterilmektedir.

Çizelge 4.7 : Eđitim, dođrulama ve test veri seti dađılımları

	<i>n</i>	%
Eđitim	34,108	68
Dođrulama	6,020	12
Test	10,032	20
Toplam	50,160	100

Bu nihai veri setinin duygu sınıfı ile dađılımı da Çizelge 4.8’de gösterilmektedir.

Çizelge 4.8 : Eđitim, dođrulama ve test veri seti ile duygu sınıfı dađılımları

Duygu Sınıfı	Eđitim		Dođrulama		Test		Toplam	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Pozitif	27,167	79.65	4,795	79.65	7,995	79.65	39,957	79.65
Nötr	507	1.49	90	1.5	150	1.5	747	1.49
Negatif	6,434	18.86	1,135	18.85	1,887	18.85	9,456	18.86
Toplam	34,108	100	6,020	100	10,032	100	50,160	100

Bu yapı sabitlendikten sonra, hem klasik makine öğrenmesi yaklaşımlarında hem de transformer tabanlı derin öğrenme modellerinde ortak kullanılmak üzere tek ve statik bir veri ayırımı yapılmıştır. Böylece tüm deneylerde aynı eğitim/dođrulama/test örnekleri kullanılarak modeller arası kıyaslanmanın adil ve tekrarlanabilir olması sağlanmıştır.

Veri kümesindeki toplam 50.160 yorum, sınıf dađılımı korunacak şekilde üçe bölünmüştür. İlk aşamada, `train_test_split` fonksiyonu stratify parametresi ile kullanılarak yaklaşık %68 eğitim (34.108 örnek) ve %20 test (10.032 örnek) kümesi elde edilmiş, kalan kısım ise eğitim kümesi üzerinden ikinci bir bölme ile %12 dođrulama (6.020 örnek) olarak ayrılmıştır. Bu dengeli ayırım sonucunda her bir alt kümede sınıf oranlarının korunduđu görülmüştür: negatif sınıf yaklaşık %18,9, nötr sınıf yaklaşık %1,5 ve pozitif sınıf yaklaşık %79,6 oranında temsil edilmiştir. Etiketler, modellerle uyumlu olacak şekilde negatif=0, nötr=1, pozitif=2 biçiminde tamsayıya çevrilmiş ve `label2id / id2label` eşleştirmeleri bütün deneylerde ortak kullanılmıştır. Aynı rastgele tohum (`seed=42`) kullanılarak ayırımın deterministik olması ve sonuçların tekrar üretilebilirliđi güvence altına alınmıştır.

Bu üçlü ayırım (eđitim / dođrulama / test), tüm temel modeller için ortak olacak şekilde sabit tutulmuştur. Ayrıca, özellikle hiperparametre optimizasyonu ve çok sayıda deneme gerektiren deneyler için, eğitim ve dođrulama kümelerinin sınıf dağılımını koruyan orantılı ama daha küçük bir alt kümesi oluşturulmuştur ve Çizelge 4.9’da gösterilmektedir.

Çizelge 4.9 : Optuna optimizasyonu denemelerinde kullanılan alt veri seti

Duygu Sınıfı	Eđitim		Dođrulama		Toplam	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Pozitif	15,930	79.65	2,788	79.66	18,718	79.65
Nötr	297	1.49	52	1.49	349	1.49
Negatif	3,773	18.86	660	18.86	4,433	18.86
Toplam	20,000	100	3,500	100	23,500	100

Sonuç olarak eğitim kümesi 20.000 örneđe, dođrulama kümesi ise 3.500 örneđe indirgenmiş; ancak her iki alt kümede de negatif, nötr ve pozitif sınıfların oransal dağılımı ana veri kümesi ile neredeyse birebir korunmuştur. Nötr sayısının az olması sebebiyle makul düzey olan minimum 35 adet yorumu koruyacak şekilde sınıf boyutuna karar verilmiştir. Bu yaklaşım sayesinde, hem hesaplama maliyeti düşürülmüş hem de azınlık sınıf olan nötr etiketin dođrulama ve eğitim setlerinde yeterli sayıda temsil edilmesi sağlanmıştır.

4.5.3 Temel model araştırmaları

Bu çalışma kapsamında duygu analizi için ince ayar sürecine geçmeden önce, kullanılacak temel modellerin (base models) belirlenmesi amacıyla kapsamlı bir literatür taraması ve model deđerlendirmesi yapılmıştır. İnceleme sürecinde modellerin mimari yapıları, eğitim stratejileri, büyük–küçük harf duyarlılığı (cased/uncased), Türkçe dil desteđi, sözlük boyutu, birimleyici (tokenizer), güncelliđi, popülerlik düzeyi, ve sosyal medya/e-ticaret diliyle uyumu gibi kriterler dikkate alınmıştır.

Bu kapsamda BERT, RoBERTa, ELECTRA, ModernBERT ve DeBERTa-V3 gibi farklı transformer mimarileri incelenmiş, duygu analizi üzerindeki etkileri deđerlendirilmiştir. Özellikle Türkçe gibi eklemeli yapıya sahip bir dilde, model mimarisinin morfolojik duyarlılığı ile birimleyici (tokenizer) yöntemi ve boyutları kritik öneme sahip olduğundan, modeller bu açılardan karşılaştırılarak seçilmiştir.

Yapılan araştırma sonucunda, hem literatürde yaygın olarak kullanılan hem de Türkçe dil işleme çalışmalarında güçlü performans sergilediği bilinen modeller listelenmiş; aynı zamanda güncel, büyük veri üzerinde eğitilmiş, ancak duygu analizi özelinde ince ayar yapılmamış modeller özellikle tercih edilmiştir.

Yürütülen değerlendirmeler sonucunda ince ayar aşamasında kullanılmak üzere aşağıdaki altı temel model seçilmiştir (dbmdz, 2020a, 2020b; VRLLab, 2023; Trendyol, 2022; Microsoft, 2021; Artiwise AI, 2023).

- dbmdz/electra-base-turkish-mc4-uncased-discriminator
- dbmdz/bert-base-turkish-128k-uncased
- VRLLab/TurkishBERTweet
- Trendyol/tyroberta
- microsoft/mdeberta-v3-base
- artiwise-ai/modernbert-base-tr-uncased

Çizelge 4.10’de görüldüğü üzere bu modeller, hem mimari çeşitlilik sunmaları hem de Türkçe derlem üzerinde eğitilmiş olmaları ya da Türkçe ile uyumlu birimleyici–sözlük yapısına sahip olmaları nedeniyle duygu analizi problemleri için güçlü bir temel oluşturmaktadır. Modeller arasında modern ön eğitim stratejilerini kullanan ModernBERT ve DeBERTa-V3, sosyal medya odaklı dil kullanımlarına yönelik olarak geliştirilen BERTweet, e-ticaret alanına özgü metinler üzerinde eğitilmiş TyRoBERTa ve Türkçe dil yapısına özel olarak optimize edilmiş DBMDZ modellerinin yer alması, farklı metin türleri ve kullanım bağlamları için model genelleme yeteneğinin karşılaştırmalı biçimde incelenmesine olanak tanımaktadır.

Tüm modeller, Transformer encoder mimarisi üzerine inşa edilmiş olup, duygu analizi gibi sınıflandırma görevleri için ek bir sınıflandırma katmanı aracılığıyla kolaylıkla ince ayar yapılabilecek şekilde tasarlanmıştır. Yaklaşık 80–135 milyon parametre aralığında değişen yönetilebilir model boyutları, hem GPU bellek kısıtları hem de eğitim süresi açısından tez kapsamındaki deneylerin sürdürülebilir ve tekrarlanabilir biçimde gerçekleştirilmesine imkân tanımaktadır. Bu yönüyle seçilen modeller, performans, kaynak kullanımı ve uygulama pratikliği arasında dengeli bir karşılaştırma zemini sunmaktadır.

Çizelge 4.10 : Temel modeller

Model ID	Mimari / Tip	Harf Duyarlılığı	Dil(ler)	Ön-Eğitim Verisi / Alanı
artwise-ai/modernbert-base-tr-uncased	ModernBERT	Duyarsuz	Türkçe	CulturaX Türkçe alt kümesi (~192 GB web metni) üzerinden, ön-eğitim
Trendyol/tyroberta	RoBERTa	Duyarlı	Türkçe	Ağırlıklı Trendyol ürün yorumları/ürün metinleri; Türkçe e-ticaret yorum korpusu üzerinde yeniden ön-eğitim yapılmış bir model olarak tanımlanıyor.
Dbmdz/electra-base-turkish-mc4-uncased-discriminator	ELECTRA	Duyarsuz	Türkçe	BERTurk ile aynı Türkçe korpus üzerinde (OSCAR, Wikipedia, OPUS, çeşitli web metinleri) ile ön-eğitim yapılmış.
VRLLab/TurkishBERTweet	BERTweet / RoBERTa	Duyarlı	Türkçe	Türkçe Twitter verisi (milyonlarca tweet) üzerinden, kısa, gürültülü, sosyal medya diline yönelik ön-eğitim.
Dbmdz/bert-base-turkish-128k-uncased	BERT	Duyarsuz	Türkçe	≈35 GB'lık Türkçe korpus: OSCAR, Wikipedia, OPUS, çeşitli haber ve kitap kaynakları;
microsoft/mdeberta-v3	DeBERTaV3	Duyarlı	Çok Dil	CC100 ve benzeri Türkçe dahil çok dilli web korpusları üzerinde,

4.5.4 Varsayılan parametreler ile eğitimler

Her bir temel model (ELECTRA, BERT, TurkishBERTweet, TyRoBERTa, mDeBERTa, ModernBERT) için varsayılan hiperparametre değerleri kullanılarak tam ince ayar gerçekleştirilmiş olup temel süreç:

- Veri setinin yüklenmesi,
- Temel modelin yüklenmesi,
- Varsayılan hiperparametre değerleri ile final ince ayar,
- Eğitilen modelin yerel bilgisayara yüklenmesi,
- Sonuçların doğrulama ve test veri setleri ile değerlendirilmesi ve raporlanması,
- Modelin Hugging Face ortamında paylaşılması,

şeklindedir.

Her bir model için eğitim, yalnızca eğitim ve doğrulama veri kümeleri üzerinde gerçekleştirilmiştir. Model karşılaştırmaları doğrulama kümesi üzerinden yapılmış; en sonunda da nihai performans değerlendirmesi yalnızca ayrılmış bağımsız test kümesi üzerinde raporlanmıştır.

Bu süreçte, ön işleme ve filtreleme adımları sonrasında elde edilen tam nihai veri setleri kullanılmış olup, ilgili varsayılan hiperparametre değerleri aşağıda sunulmaktadır:

```
params = {  
  "learning_rate": 5e-05,  
  "per_device_train_batch_size": 8,  
  "per_device_eval_batch_size": 8,  
  "weight_decay": 0.0,  
  "warmup_ratio": 0.0,  
  "num_train_epochs": 3,  
  "gradient_accumulation_steps": 1  
}
```

Belirlenen bu hiperparametre yapılandırması, literatürde yaygın olarak kullanılan varsayılan ayarlar ile uyumlu olup, farklı model mimarileri arasında adil ve karşılaştırılabilir bir değerlendirme yapılmasına olanak tanımaktadır.

4.5.5 Optuna parametre optimizasyonu ile eğitimler

Her bir model için eğitim, yalnızca eğitim kümesi üzerinde gerçekleştirilmiş; hiperparametre seçimleri ve model karşılaştırmaları doğrulama kümesi üzerinden yapılmış; en sonunda da nihai performans değerlendirmesi bağımsız test kümesi üzerinde raporlanmıştır. Hiperparametre optimizasyonu Optuna ile yürütülmüş olup temel süreç:

- Veri setinin yüklenmesi,
- Temel modelin yüklenmesi,
- Hiperparametre optimizasyon aralıklarının ayarlanması,
- Alt veri seti kullanılarak hiperparametre ince ayar denemeleri,
- Tam veri seti kullanılarak en iyi hiperparametre değerleri ile final ince ayar,
- Eğitilen modelin yerel bilgisayara yüklenmesi,
- Sonuçların doğrulama ve test veri setleri ile değerlendirilmesi ve raporlanması,
- Modelin Hugging Face ortamında paylaşılması,

şeklinde.

Bu süreçte, tam veri kümesi yerine önceki adımda oransal olarak küçültülmüş ve sınıf dağılımı korunmuş 23,500 yorum içeren eğitim ve doğrulama veri setleri kullanılmıştır. İlgili modelin tokenizer'ı da yüklendikten sonra, yorum metinleri maksimum 256 token olacak şekilde yalnızca kırpma (truncation) uygulanarak tokenleştirilmiştir. Sabit bir DataCollatorWithPadding nesnesi ile dinamik padding kullanılmış, böylece her yığın (batch) içinde sekans uzunluklarına göre otomatik doldurma yapılarak GPU belleği daha verimli kullanılmıştır.

```
params = {
    "learning_rate": trial.suggest_float("learning_rate", 5e-6, 5e-5, log=True),
    "per_device_train_batch_size": trial.suggest_categorical("per_device_train_batch_size", [16, 32]),
    "per_device_eval_batch_size": trial.suggest_categorical("per_device_eval_batch_size", [32]),
    "weight_decay": trial.suggest_float("weight_decay", 0.0, 0.1),
    "warmup_ratio": trial.suggest_float("warmup_ratio", 0.0, 0.2),
    "num_train_epochs": trial.suggest_int("num_train_epochs", 6, 8),
    "gradient_accumulation_steps": trial.suggest_categorical("gradient_accumulation_steps", [1, 2, 4]),
}
```

Optimizasyonun merkezinde, Optuna'ya tanımlanan objective(trial) fonksiyonu yer almaktadır. Her bir denemede hiperparametre aralıkları yukarıda tanımlandığı gibi ayarlanmıştır. Bu parametreler, temel TrainingArguments yapısının üzerine yazılarak her deneme için ayrı bir eğitim konfigürasyonu oluşturulmuştur. Model, model_init fonksiyonu ile her deneme başında sıfırdan yüklenmiş, böylece denemeler arası ağırlık taşması engellenmiştir.

Değerlendirme tarafında, Hugging Face Trainer yapısına özel bir compute_metrics fonksiyonu entegre edilmiştir. Bu fonksiyon, doğrulama kümesi tahminleri üzerinden doğruluk, dengeli doğruluk, Matthews korelasyon katsayısı (MCC) ve makro/mikro/ağırlıklı F1, kesinlik ve duyarlılık değerlerini hesaplamaktadır. Ayrıca her sınıf (pozitif, nötr, negatif) için ayrı ayrı kesinlik/duyarlılık/F1 metrikleri üretilmekte ve class_pozitif_f1 vb. anahtarlarla kaydedilmektedir. Optuna hedef metrik eval_macro_f1 olarak tanımlanmıştır. Böylece çok sınıflı ve dengesiz veri yapısına daha duyarlı bir optimizasyon kriteri elde edilmiştir.

Eğitim süreci boyunca hem erken sonlandırma (pruning) hem de ayrıntılı loglama mekanizmaları devreye alınmıştır. OptunaPruningCallbackHF sınıfı, her doğrulama adımında eval_macro_f1 değerini Optuna'ya rapor etmekte; median-pruner stratejisi uyarınca, benzer denemelere kıyasla belirgin biçimde geride kalan denemeler erken aşamada trial pruned hatası ile sonlandırılmaktadır. Böylece özellikle 6–8 epoch aralığında çalışan denemelerde süre ve GPU kaynağı tasarrufu sağlanmıştır. Paralel olarak, her bir deneme ve denemelerdeki epoch ile hesaplanan değerler ayrıca CSV olarak raporlanmıştır. Bu CSV çıktıları sayesinde, 100 denemenin her biri için epoch düzeyinde performans eğrileri ve hiperparametre-performans ilişkisi ayrıntılı biçimde analiz edilebilmiştir.

Her temel model için toplam 100 Optuna denemesi yürütülmüştür. Optuna çalışması tamamlandığında, en yüksek hedef değere (makro F1) sahip en iyi deneme otomatik olarak seçilmiş ve ilgili model-hiperparametre kombinasyonu, daha sonra tam eğitim ve doğrulama veri setleri ile son ince ayar eğitimine tabi tutulmuştur.

Bu süreçle ilgili her bir deneme düzeyindeki örnek epoch çıktıları için Çizelge 4.11 bakılabilir. Bütün denemeler sonucunda en iyi denemeyi veren örnek için ise Çizelge 4.12 bakılabilir.

Çizelge 4.11 : Optuna HPO denemeleri epoch düzeyi

Deneme	Epoch	Eğitim Kaybı	Doğrulama Kaybı	Doğruluk	Dengelenmiş Doğruluk	Makro F1 - Hedef Metrik	Kesinlik	Duyarlılık
...
21	1	0.2787	0.1777	0.9360	0.6126	0.6049	0.5976	0.6126
21	2	0.1638	0.1811	0.9349	0.6013	0.6025	0.6040	0.6013
21	3	0.1243	0.1960	0.9346	0.6767	0.6905	0.7077	0.6767
21	4	0.0906	0.2806	0.9334	0.6203	0.6315	0.6658	0.6203
21	5	0.0615	0.3744	0.9277	0.6058	0.6069	0.6088	0.6058
21	6	0.0412	0.4278	0.9266	0.6426	0.6439	0.6453	0.6426
21	7	0.0291	0.4393	0.9249	0.6600	0.6520	0.6477	0.6600
21	8	0.0197	0.4539	0.9257	0.6510	0.6491	0.6494	0.6510
...

Çizelge 4.12 : Optuna HPO denemeleri

Deneme	Makro F1 - Hedef Metrik	Öğrenme Oranı	Eğitim Yığın Boyutu	Doğrulama Yığın Boyutu	Ağırlık Azalışı	Isınma Oranı	Eğitim Epoch Sayısı	Gradyan Biriktirme Adım Sayısı
0	0.62215088	0.00000628	16	32	0.00914280	0.09255109	6	2
1	0.65252266	0.00004570	16	32	0.09797843	0.08787411	8	2
2	0.64691537	0.00003669	32	32	0.08963125	0.17636103	8	1
3	0.64610533	0.00001251	32	32	0.06390634	0.09698584	8	1
4	0.62159183	0.00000608	32	32	0.03266559	0.05062150	7	1
5	0.60309492	0.00000599	32	32	0.01478619	0.12111060	6	2
6	0.58979359	0.00003656	16	32	0.07437597	0.08302229	8	2
7	0.60168848	0.00002893	16	32	0.07119244	0.06482719	8	1
8	0.58721572	0.00004162	32	32	0.07272403	0.12631098	7	1
9	0.64289861	0.00003322	32	32	0.06784303	0.08677681	6	2
10	0.60197029	0.00001977	16	32	0.09909628	0.00181846	7	4
11	0.63549425	0.00004947	16	32	0.09897804	0.19764365	8	4
12	0.59843021	0.00002434	16	32	0.08518383	0.18278092	8	2
13	0.60081287	0.00001287	32	32	0.04754166	0.15792044	8	1
14	0.64395159	0.00004889	32	32	0.08770213	0.14387394	7	4
15	0.60235480	0.00002043	16	32	0.05079164	0.04779041	7	2
16	0.65187196	0.00002799	32	32	0.08729766	0.16568080	8	1
17	0.65629145	0.00002654	16	32	0.05440719	0.12204237	8	2
18	0.65524425	0.00001479	32	32	0.08243905	0.01978435	8	1
19	0.60340101	0.00001023	16	32	0.03722743	0.00368883	7	4
...

4.6 Sonuçların Değerlendirilmesi

Model eğitimlerinin tamamlanması ile hem doğrulama hem de test veri setleri üzerinden modellerin performansları genel ve sınıf bazlı metrikler üzerinden değerlendirmeye tabi tutulmuştur. Doğrulama ve test kümesi sonuçlarının birlikte ele alınması, modellerin yalnızca eğitim verisine uyum sağlama düzeylerini değil, aynı zamanda daha önce görülmemiş veriler üzerindeki genelleme yeteneklerini de değerlendirmeye olanak tanımaktadır. Özellikle doğrulama ve test performansları arasındaki farklar incelenerek, aşırı uyum eğilimleri ve model kararlılığı analiz edilmiştir. Sınıf bazlı metrikler üzerinden yapılan değerlendirmeler ise pozitif, negatif ve nötr sınıflar arasındaki performans dengesini ortaya koyarak, modellerin hangi duygu sınıflarında daha güçlü veya zayıf kaldığını görünür kılmaktadır.

İlgili sonuçlar, hem Optuna tabanlı hiperparametre optimizasyonu uygulanmış ince ayarlı modelleri hem de varsayılan (hiperparametre optimizasyonu uygulanmamış) modelleri kapsamaktadır. Bu sayede, hiperparametre optimizasyonunun model performansı üzerindeki etkisi doğrudan karşılaştırmalı olarak analiz edilebilmekte ve elde edilen kazanımların tutarlılığı değerlendirilebilmektedir.

Dengesiz sınıf dağılımına sahip veri setlerinde Makro-F1, modellerin azınlık sınıfları öğrenme performansını daha iyi yansıtan bir ölçüt olduğu için eğitim sürecinde bu değerlerin maksimize edilmesi amaçlanmış olup, karşılaştırmalar bu metrik üzerinden yapılmaktadır.

Doğrulama Veri Seti üzerinden Genel Metrikler ile Analiz (Çizelge 4.13):

- En yüksek Makro-F1: değeri 0.6548 ile Optuna ile hiperparametre optimizasyonu yapılan Trendyol/Tyroberta modeli ile gerçekleşmiştir.
- En yüksek Kesinlik: değeri 0.6566 ile Optuna ile hiperparametre optimizasyonu yapılan microsoft/mdeberta-v3 modeli ile gerçekleşmiştir.
- En yüksek Duyarlılık: değeri 0.6635 ile Optuna ile hiperparametre optimizasyonu yapılan Trendyol/tyroberta modeli ile gerçekleşmiştir.
- Optuna ile hiperparametre optimizasyonu uygulanmış modeller, uygulanmamış modellerin yaklaşık %4–7 daha yüksek Makro-F1 performansı göstermektedir. Bu fark, ince ayarın ve özellikle hiperparametre

optimizasyonunun dengesiz veri üzerinde kritik bir performans iyileştiricisi olduğunu göstermektedir.

- Optuna uygulanan her modelde Makro-F1 belirgin şekilde artmış ve en kötü Optuna ile hiperparametre optimizasyonu sonucu bile diğer varsayılan modellerden daha iyi performans göstermiştir.

Test Veri Seti üzerinden Genel Merikler ile Analiz (Çizelge 4.14):

- En yüksek Makro-F1: değeri 0.6417 ile Optuna ile hiperparametre optimizasyonu uygulanmış dbmdz/bert modelinden elde edilmiştir.
- En yüksek Kesinlik: değer 0.6449 ile yine Optuna ile optimize edilmiş dbmdz/bert modelinden elde edilmiştir.
- En yüksek Duyarlılık: değeri 0.6423 ile Optuna tarafından optimize edilen Trendyol/tyroberta modelinden elde edilmiştir.
- Optuna ile hiperparametre optimizasyonu uygulanmış modeller, varsayılan modellere kıyasla yaklaşık %3–5 oranında daha yüksek Makro-F1 performansı sergilemiş olup, bu fark hiperparametre optimizasyonunun dengesiz veri koşullarında model performansını belirgin biçimde iyileştirdiğini göstermektedir.

Doğrulama Veri Seti üzerinden Sınıf Bazlı Merikler ile Analiz (Çizelge 4.15):

- Optuna ile optimize edilen modeller, negatif ve pozitif sınıflarda varsayılan modellere kıyasla belirgin performans artışı sağlamıştır.
- Nötr sınıftaki düşük başarı, veri kümesindeki sınıf dengesizliğinin ve nötr sınıfın semantik açıdan daha belirsiz yapısının doğal bir sonucudur.
- Optuna uygulaması, Kesinlik ve Duyarlılık değerlerinin uyumlu şekilde yükselmesini sağlayarak modellerin genel sınıf bazlı performansını daha dengeli hâle getirmiştir.
- Negatif Sınıf
 - Negatif sınıfı için en yüksek F1 skoru, Optuna ile optimize edilen Trendyol/Tyroberta modelinde elde edilmiştir (F1 = 0.8357).
 - Negatif sınıfta en yüksek Kesinlik değeri (0.8465), Optuna ModernBERT modelinde görülmüştür; bu bulgu modelin negatif sınıf için daha az yanlış pozitif tahminde bulunduğunu göstermektedir.

- En yüksek Duyarlılık değeri varsayılan Trendyol/Tyroberta modeline aittir (0.8731), ancak Kesinlik daha düşük olduğundan bu model sınıf bazında dengeli bir performans sergilememektedir.
- Nötr Sınıf
 - Nötr sınıfta tüm modellerin performansı genel olarak düşük seyretmektedir; bu durum nötr sınıfın veri sayısının azlığı, sınıf sınırlarının belirsizliği ve veri dengesizliği nedeniyle öğrenilmesinin zor olduğunu göstermektedir.
 - Nötr sınıfta en yüksek F1 değeri, Optuna dbmdz/bert modelinde elde edilmiştir (F1 = 0.1900).
 - Optuna uygulaması nötr sınıftaki performansı artırsa da, nötr sınıf modeller için en zayıf performans gösterilen sınıf olmaya devam etmektedir.
- Pozitif Sınıf
 - Pozitif sınıfta en yüksek F1 skoru, Optuna Trendyol/Tyroberta modelinde elde edilmiştir (F1 = 0.9634).
 - Pozitif sınıfta en yüksek Kesinlik, yine Optuna Trendyol/Tyroberta modeline aittir (0.9654); bu model pozitif sınıfa ilişkin tahminlerde oldukça tutarlı davranmaktadır.
 - En yüksek Duyarlılık ise Varsayılan VRLLab/TurkishBERTweet modelinde görülmektedir (0.9858), bu model pozitif örnekleri yakalama konusunda en güçlü performansa sahiptir.
 - Genel Değerlendirme

Test Veri Seti üzerinden Sınıf Bazlı Merikler ile Analiz (Çizelge 4.16):

- Optuna ile optimize edilen modeller, negatif ve pozitif sınıflarda varsayılan modellere göre anlamlı performans artışı sağlamıştır.
- Nötr sınıfın düşük F1 sonuçları, nötr sınıfın gelecek çalışmalarda veri zenginleştirme gerektirdiğini işaret etmektedir.
- Optuna ile yapılan ince ayar, test veri setinde de Kesinlik ve Duyarlılık arasındaki dengeyi iyileştirmiş, modellerin sınıflar arası ayrımı daha dengeli öğrenmesini sağlamıştır.
- Negatif Sınıf

- Negatif sınıfta en yüksek F1 skoru, Optuna ile optimize edilen Trendyol/Tyroberta modelinde elde edilmiştir ($F1 = 0.8345$). Bu sonuç, Trendyol/Tyroberta modelinin test verisinde negatif sınıfı en başarılı genelleyen model olduğunu göstermektedir.
- En yüksek Kesinlik (0.8408), Optuna Microsoft/mdeberta-v3 modelinde görülmektedir; bu, modelin negatif sınıf için daha az yanlış pozitif ürettiğini göstermektedir.
- Negatif sınıfta en yüksek Duyarlılık, varsayılan Trendyol/Tyroberta modeline aittir (Duyarlılık = 0.8679). Varsayılan modelin negatif örnekleri yakalamada güçlü olduğunu, ancak Kesinlik dengesinin düşük olması nedeniyle genel performansın dengeli olmadığını göstermektedir.
- Nötr Sınıf
 - Nötr sınıf performansı tüm modellerde düşük seviyededir; bu durum sınıfın veri azlığı, belirsiz semantik yapısı ve sınıf dengesizliği nedeniyle öğrenilmesinin zor olduğunu göstermektedir.
 - Nötr sınıfta en yüksek F1 skorunu Optuna dbmdz/bert modeli üretmiştir ($F1 = 0.1597$). Bu değer düşük olmakla birlikte tüm modeller arasında en yüksek başarımlar olarak öne çıkmaktadır.
 - Optuna ile optimize edilmiş modeller, nötr sınıfta varsayılan modellere göre daha yüksek F1 üretmiş olsa da, nötr sınıf yine en zayıf öğrenilen sınıf olarak kalmıştır.
- Pozitif Sınıf
 - Pozitif sınıfta en yüksek F1 skoru, Optuna Trendyol/Tyroberta modeline aittir ($F1 = 0.9598$).
 - En yüksek Duyarlılık değeri (0.9865), varsayılan VRLLab/TurkishBERTtweet modelinde görülmüştür. Bu sonuç, bu modelin pozitif sınıf tahminlerine yönelik hata oranının düşük olduğunu göstermektedir. Ama kesinlik değerinin düşük olması sebebiyle genel performansı düşük kalmaktadır.
 - Pozitif sınıfta en yüksek Kesinlik, Optuna Trendyol/Tyroberta modelinde elde edilmiştir (Kesinlik 0.9633). Bu, modelin pozitif örnekleri yakalama açısından en güçlü performansı verdiğini göstermektedir.

Çizelge 4.13 : Doğrulama veri seti ile önerilen modellerin performans analizi

Tip	İnce Ayarlı Model	Doğruluk	F1 (Makro)	Kesinlik (Makro)	Duyarlılık (Makro)	Destek (N)
Optuna	Trendyol/tyroberta	0.9251	0.6548	0.6495	0.6635	6020
Optuna	dbmdz/bert	0.9193	0.6530	0.6559	0.6521	6020
Optuna	dbmdz/electra	0.9208	0.6420	0.6515	0.6334	6020
Optuna	microsoft/mdeberta-v3	0.9236	0.6383	0.6566	0.6257	6020
Optuna	artwise-ai/modernbert	0.9216	0.6376	0.6419	0.6345	6020
Varsayılan	artwise-ai/modernbert	0.9184	0.6317	0.6362	0.6284	6020
Optuna	VRLLab/TurkishBERTweet	0.9140	0.6305	0.6358	0.6265	6020
Varsayılan	dbmdz/electra	0.9309	0.5990	0.5935	0.6046	6020
Varsayılan	microsoft/mdeberta-v3	0.9299	0.5986	0.5906	0.6070	6020
Varsayılan	Trendyol/tyroberta	0.9291	0.5986	0.5873	0.6110	6020
Varsayılan	dbmdz/bert	0.9173	0.5848	0.5795	0.5901	6020
Varsayılan	VRLLab/TurkishBERTweet	0.8236	0.4066	0.5259	0.3964	6020
Varsayılan	Tamamlayıcı Naif Bayes	0.8626	0.6087	0.5989	0.6921	6020
Varsayılan	K-En Yakın Komşuluk	0.8782	0.5548	0.5863	0.5359	6020

Çizelge 4.14 : Test veri seti ile önerilen modellerin performans analizi

Tip	İnce Ayarlı Model	Doğruluk	F1 (Makro)	Kesinlik (Makro)	Duyarlılık (Makro)	Destek (N)
Optuna	dbmdz/bert	0.9169	0.6417	0.6449	0.6391	10032
Optuna	Trendyol/tyroberta	0.9215	0.6364	0.6322	0.6423	10032
Optuna	dbmdz/electra	0.9192	0.6338	0.6404	0.6279	10032
Varsayılan	artwise-ai/modernbert	0.9202	0.6319	0.6356	0.6286	10032
Optuna	artwise-ai/modernbert	0.9173	0.6282	0.6323	0.6255	10032
Optuna	microsoft/mdeberta-v3	0.9221	0.6264	0.6419	0.6160	10032
Optuna	VRLLab/TurkishBERTweet	0.9117	0.6220	0.6229	0.6216	10032
Varsayılan	dbmdz/electra	0.9272	0.5958	0.5909	0.6008	10032
Varsayılan	Trendyol/tyroberta	0.9254	0.5956	0.5843	0.6081	10032
Varsayılan	microsoft/mdeberta-v3	0.9242	0.5940	0.5847	0.6040	10032
Varsayılan	dbmdz/bert	0.9200	0.5893	0.5830	0.5957	10032
Varsayılan	VRLLab/TurkishBERTweet	0.8230	0.4044	0.5312	0.3947	10032
Varsayılan	Tamamlayıcı Naif Bayes	0.8543	0.5942	0.5891	0.6628	10032
Varsayılan	K-En Yakın Komşuluk	0.8753	0.5594	0.5850	0.5416	10032

Çizelge 4.15 : Doğrulama veri seti ile önerilen modellerin sınıf bazlı performans analizi

Tip	Model \ Sınıf İnce Ayarlı Model	Negatif			Nötr			Pozitif		
		Kesinlik	Duyarlılık	F1	Kesinlik	Duyarlılık	F1	Kesinlik	Duyarlılık	F1
Optuna	microsoft/mdeberta-v3	0.8411	0.7974	0.8186	0.1786	0.1111	0.1370	0.9503	0.9687	0.9594
Optuna	dbmdz/electra	0.8366	0.7894	0.8123	0.1667	0.1444	0.1548	0.9513	0.9664	0.9588
Optuna	dbmdz/bert	0.8421	0.7797	0.8097	0.1727	0.2111	0.1900	0.9529	0.9656	0.9592
Optuna	artwise-ai/modernbert	0.8465	0.7921	0.8184	0.1238	0.1444	0.1333	0.9553	0.9668	0.9610
Optuna	VRLLab/TurkishBERTweet	0.8004	0.8018	0.8011	0.1549	0.1222	0.1366	0.9520	0.9554	0.9537
Optuna	Trendyol/tyroberta	0.8424	0.8291	0.8357	0.1406	0.2000	0.1651	0.9654	0.9614	0.9634
Varsayılan	microsoft/mdeberta-v3	0.8127	0.8564	0.8340	0.0000	0.0000	0.0000	0.9590	0.9648	0.9618
Varsayılan	dbmdz/electra	0.8239	0.8449	0.8343	0.0000	0.0000	0.0000	0.9565	0.9687	0.9626
Varsayılan	dbmdz/bert	0.7911	0.8106	0.8007	0.0000	0.0000	0.0000	0.9475	0.9597	0.9536
Varsayılan	artwise-ai/modernbert	0.8418	0.7877	0.8138	0.1143	0.1333	0.1231	0.9526	0.9641	0.9583
Varsayılan	VRLLab/TurkishBERTweet	0.7500	0.2035	0.3202	0.0000	0.0000	0.0000	0.8276	0.9858	0.8998
Varsayılan	Trendyol/tyroberta	0.7992	0.8731	0.8345	0.0000	0.0000	0.0000	0.9628	0.9597	0.9613
Varsayılan	Tamamlayıcı Naif Bayes	0.7382	0.7727	0.7551	0.0841	0.4111	0.1396	0.9743	0.8924	0.9315
Varsayılan	K-En Yakın Komşuluk	0.7547	0.6044	0.6712	0.0976	0.0444	0.0611	0.9067	0.9587	0.9320

Çizelge 4.16 : Test veri seti ile önerilen modellerin sınıf bazlı performans analizi

Tip	Model \ Sınıf İnce Ayarlı Model	Negatif			Nötr			Pozitif		
		Kesinlik	Duyarlılık	F1	Kesinlik	Duyarlılık	F1	Kesinlik	Duyarlılık	F1
Optuna	microsoft/mdeberta-v3	0.8408	0.8008	0.8203	0.1364	0.0805	0.1013	0.9487	0.9666	0.9575
Optuna	dbmdz/electra	0.8284	0.8008	0.8144	0.1406	0.1208	0.1300	0.9521	0.9621	0.9570
Optuna	dbmdz/bert	0.8307	0.7882	0.8089	0.1524	0.1678	0.1597	0.9516	0.9613	0.9564
Optuna	artwise-ai/modernbert	0.8393	0.7861	0.8118	0.1050	0.1275	0.1152	0.9526	0.9631	0.9578
Optuna	VRLLab/TurkishBERTweet	0.7922	0.8056	0.7988	0.1231	0.1074	0.1147	0.9534	0.9518	0.9526
Optuna	Trendyol/tyroberta	0.8327	0.8362	0.8345	0.1005	0.1342	0.1149	0.9633	0.9563	0.9598
Varsayılan	microsoft/mdeberta-v3	0.7980	0.8537	0.8249	0.0000	0.0000	0.0000	0.9562	0.9582	0.9572
Varsayılan	dbmdz/electra	0.8198	0.8362	0.8279	0.0000	0.0000	0.0000	0.9528	0.9661	0.9594
Varsayılan	dbmdz/bert	0.8000	0.8283	0.8139	0.0000	0.0000	0.0000	0.9491	0.9588	0.9539
Varsayılan	artwise-ai/modernbert	0.8401	0.8019	0.8205	0.1125	0.1208	0.1165	0.9541	0.9631	0.9586
Varsayılan	VRLLab/TurkishBERTweet	0.7680	0.1976	0.3143	0.0000	0.0000	0.0000	0.8258	0.9865	0.8990
Varsayılan	Trendyol/tyroberta	0.7930	0.8679	0.8288	0.0000	0.0000	0.0000	0.9599	0.9563	0.9581
Varsayılan	Tamamlayıcı Naif Bayes	0.7282	0.7771	0.7519	0.0610	0.3289	0.1029	0.9779	0.8824	0.9277
Varsayılan	K-En Yakın Komşuluk	0.7582	0.6112	0.6768	0.0909	0.0604	0.0726	0.9058	0.9531	0.9288

4.7 Modellerin Yayınlanması

Eğitim ve performans değerlendirmeleri sonrasında eğitilen bütün modeller kapsamlı açıklamaları ile Hugging Face ortamında yayınlanarak, gelecek araştırmacıların desteklenmesi amaçlanmıştır.

Modellerin yerel ortamda eğitilmesi sonrasında, eğitim sürecinde üretilen config.json, tokenizer.json ve model.safetensor gibi temel dosyalar en güncel mimari ile depoya yüklenmiştir. Ardından, Jupyter Notebook ortamında hazırlanan algoritmalar aracılığıyla hem model kartı hem de modele ait teknik açıklamalar, kullanım örnekleri ve eğitim metrikleri Hugging Face formatına uygun biçimde otomatik üretilip yayınlanmıştır:

- tyroberta-turkish-sentiment-optuna-hpo (msamilim, 2025a)
- mdeberta-v3-turkish-sentiment-optuna-hpo (msamilim, 2025b)
- bert-128k-turkish-sentiment-optuna-hpo (msamilim, 2025c)
- turkishbertweet-turkish-sentiment-optuna-hpo (msamilim, 2025d)
- modernbert-turkish-sentiment-optuna-hpo (msamilim, 2025e)
- electra-turkish-sentiment-optuna-hpo (msamilim, 2025f)

Bu işlem sonucunda, ince ayar yapılmış modeller hem yerel kullanım hem de bulut tabanlı uygulamalar ve araştırmacılar için erişilebilir hâle gelmiştir.

Yayınlanmış modellerden örnek olarak, “Trendyol/tyroberta” temel modeli üzerinden ince ayar yapılmış model dökümantasyon detayları aşağıda yer almaktadır:

Hiperparametre arama uzayı aşağıdaki gibidir:

```
params = {
  "learning_rate": trial.suggest_float("learning_rate", 5e-6, 5e-5, log=True),
  "per_device_train_batch_size": trial.suggest_categorical("per_device_train_batch_size", [16, 32]),
  "per_device_eval_batch_size": trial.suggest_categorical("per_device_eval_batch_size", [32]),
  "weight_decay": trial.suggest_float("weight_decay", 0.0, 0.1),
  "warmup_ratio": trial.suggest_float("warmup_ratio", 0.0, 0.2),
  "num_train_epochs": trial.suggest_int("num_train_epochs", 6, 8),
  "gradient_accumulation_steps": trial.suggest_categorical("gradient_accumulation_steps", [1, 2, 4]),
}
```

100 deneme sonunda bulunan en iyi hiperparametre değerleri aşağıdaki gibidir:

```
{
  "learning_rate": 1.6142125050800487e-05,
  "per_device_train_batch_size": 32,
  "per_device_eval_batch_size": 32,
  "weight_decay": 0.0932597392123772,
  "warmup_ratio": 0.0627483542343149,
  "num_train_epochs": 8,
  "gradient_accumulation_steps": 1
}
```

Doğrulama veri seti ile değerlendirme sonuçları Çizelge 4.17’de gösterilmektedir.

Çizelge 4.17 : Doğrulama veri seti ile önerilen modelin performans analizi

Etiket	Kesinlik	Duyarlılık	F1 Skor
negatif	0.8424	0.8291	0.8357
nötr	0.1406	0.2000	0.1651
pozitif	0.9654	0.9614	0.9634
doğruluk			0.9251
micro ort	0.9251	0.9251	0.9251
makro ort	0.6495	0.6635	0.6548
ağırlıklı ort	0.9299	0.9251	0.9274

Eğitim epoch adım raporlamaları Çizelge 4.18’de gösterilmektedir.

Çizelge 4.18 : Önerilen modelin epoch adım raporları

Epoch	Eğitim Kaybı	Doğrulama Kaybı	Makro F1
1	0.2595	0.1998	0.6011
2	0.1632	0.1969	0.6058
3	0.1286	0.2104	0.6358
4	0.0959	0.2964	0.6379
5	0.0701	0.3393	0.6386
6	0.0507	0.3769	0.6244
7	0.0357	0.4328	0.6548

Hugging Face kütüphanesi aracılığıyla ilgili modelin pipeline olarak kullanımı:

```
from transformers import pipeline
# Load the classification pipeline with the specified model
model_name = "msamilim/tyroberta-turkish-sentiment-optuna-hpo"
pipe = pipeline("text-classification", model=model_name)
# Classify a new sentence
sentence = "Güzel ürün, tavsiye ederim."
result = pipe(sentence)
# Print the result
print(result)
# Example output :
# [{'label': 'pozitif', 'score': 0.9998408555984497}]
```

Hugging Face kütüphanesi aracılığıyla ilgili modelin Tokenizer ile kullanımı:

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch
model_name = "msamilim/tyroberta-turkish-sentiment-optuna-hpo"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)
def predict_sentiment(texts):
    inputs = tokenizer(texts, return_tensors="pt", truncation=True, padding=True, max_length=512)
    with torch.no_grad():
        outputs = model(**inputs)
    probabilities = torch.nn.functional.softmax(outputs.logits, dim=-1)
    id2label = { 0: "Negatif", 1: "Nötr", 2: "Pozitif" }
    return [id2label[p] for p in torch.argmax(probabilities, dim=-1).tolist()]
texts = [
    "Güzel ürün, tavsiye ederim kullanılır.",
    "Ürün çok güzel ve kaliteli. Maalesef yüzüme uymadığı için iade etmek zorunda kaldım.",
    "Keşke aldıktan sonra indirim girmeseydi.",
    "Daha soluk ve mat yapısı var beğenmedim .",
]
for text, sentiment in zip(texts, predict_sentiment(texts)):
    print(f"Text: {text}\nSentiment: {sentiment}\n")
# Example output :
# Text: Güzel ürün, tavsiye ederim kullanılır.
# Sentiment: Pozitif
# Text: Ürün çok güzel ve kaliteli. Maalesef yüzüme uymadığı için iade etmek zorunda kaldım.
# Sentiment: Pozitif
# Text: Keşke aldıktan sonra indirim girmeseydi.
# Sentiment: Negatif
# Text: Daha soluk ve mat yapısı var beğenmedim .
# Sentiment: Negatif
```

Kullanılan mimariler:

```
transformers: 4.57.0
torch: 2.8.0+cu128
datasets: 4.2.0
accelerate: 1.10.1
evaluate: 0.4.6
python: 3.11.1
```

5. SONUÇLAR

Bu tez çalışmasında, Türkçe e-ticaret yorumlarından oluşan dengesiz bir veri seti üzerinde Transformer tabanlı modellerin performansı incelenmiş; Optuna ile hiperparametre optimizasyonu uygulanmış ince ayarlı modeller ile varsayılan modeller karşılaştırılmıştır. Çalışma kapsamında, önerilen modellerde hem doğrulama hem de test veri setleri kullanılarak genel metrikler (Doğruluk, Makro-F1, Kesinlik, Duyarlılık) ve sınıf bazlı metrikler üzerinden değerlendirmeye tabi tutulmuştur. Elde edilen sonuçlar aşağıda özetlenmektedir:

- Test veri setinde en yüksek performansın Optuna ile optimize edilen tam ince ayarlı dbmdz/bert modeli tarafından elde edildiğini göstermektedir (Macro-F1: %64,17; Doğruluk: %91,69). Doğrulama veri setinde ise Optuna ile optimize edilen tam ince ayarlı Trendyol/tyroberta en yüksek sonuçlara ulaşmıştır (Macro-F1: %65,48; Doğruluk: %92,51).
- Optuna hiperparametre optimizasyonu uygulanmış modeller, varsayılan modellere kıyasla %3–7 oranında Macro-F1 artışı sağlaması ile daha yüksek performans sergilemiştir. Hem genel hem de sınıf bazlı performansı iyileştiren kritik bir bileşen olarak öne çıkmıştır.
- Transformer bazlı modellerin klasik sınıflandırma algoritmalarına kıyasla daha yüksek Macro F1 sağladığı gözlemlenmiştir.
- Doğrulama ve test sonuçları arasındaki farkların düşük olması, modellerin genellenebilirliğinin yüksek olduğunu göstermektedir.
- Nötr sınıf, veri dengesizliği nedeniyle en zayıf performans gösteren kategoridir. Bu durum sınıfın veri yetersizliği, semantik belirsizlik ve dil karmaşası nedenleriyle öğrenmesinin zor olduğunu göstermektedir.

Bu çalışma kapsamında önerilen Transformer tabanlı duygu analizi modelleri, Türkçe e-ticaret yorumları gibi dilsel açıdan karmaşık ve gürültülü veri setleri üzerinde yüksek doğruluk ve tutarlılık sağlayarak, klasik sınıflandırma algoritmalarına kıyasla belirgin bir performans artışı ortaya koymuştur. Özellikle bağlamı koruyabilen dikkat

mekanizmaları sayesinde, ironi, örtük duygu ve bağlama bağı anlam deęişimleri gibi Türkçe'ye özgü zorlukların daha başarılı biçimde yakalandığı gözlemlenmiştir. Aynı zamanda, test veri setindeki genelleme kabiliyeti ile gerçek dünya verileri üzerinde güvenle kullanılacak olgunlukta çözümler sunduğunu göstermektedir.

Önerilen modeller; ürün yönetimi, müşteri deneyimi analizi ve ürün çeşitliliği planlaması gibi karar alma süreçlerine doğrudan katkı sağlayabilecek niteliktedir. Manuel inceleme gerektiren on binlerce müşteri yorumunun otomatik ve tutarlı biçimde analiz edilebilmesi, firmalara hem zaman hem de maliyet avantajı sunarken, müşteri beklentilerinin daha hızlı ve doğru biçimde tespit edilmesine imkan tanımaktadır. Özellikle duygu dağılımlarının ürün özellikleri, fiyat seviyesi ve kullanıcı profilleriyle birlikte değerlendirilmesi; pazarlama stratejilerinin, ürün iyileştirme kararlarının ve portföy optimizasyonunun sezgisel kararlar yerine veri temelli olarak şekillendirilmesini mümkün kılmaktadır. Böylelikle duygu analizi, yalnızca müşteri memnuniyetini ölçen bir araç olmanın ötesine geçerek; ürün geliştirme, müşteri ayrılma eğilimi tahmini, pazarlama stratejisi ve rekabet avantajı oluşturma süreçlerinde firmalar için stratejik bir karar destek mekanizması haline gelmektedir.

6. GELECEĞE İLİŞKİN ÖNERİLER

Mevcut tez çalışması kapsamında güncel transformer temel mimariler üzerine ince ayar yapılarak yeni modeller geliştirilmiş olup bu modeller Hugging Face ortamında yayınlanmıştır. Gelecek çalışmalarda model eğitimleri sürecinde, farklı hiperparametre optimizasyon yöntemleri kullanılarak karşılaştırma yapılması önerilmektedir. Ayrıca kullanılan veri seti dengesiz olup, daha fazla ve dengeli veri seti üzerinden yeni eğitimlerin gerçekleştirilmesi önerilmektedir. Bir diğer öneri ise soru-cevap metinleri veri setinin de çalışmaya entegre edilmesi olacaktır.

Mevcut geliştirilen modellerin kullanımı ile etiketlenmemiş veri seti üzerinden duygu tahmini yapılarak, ürün öneri sistemleri geliştirilmesinin de önü açıktır. Nitekim toplanan veri seti ürün yorumlarının ötesinde ürün bilgileri, benzer ürünler, soru-cevaplar ve çok daha fazlası ile oldukça zengindir. Bu kapsamda Şekil 6.1’de görüldüğü üzere, araştırmacılar ve uygulamacılar için örnek bir arayüz tasarımı kodlanmıştır.

Dikdörtgen formunda, benx marka güneş gözlüğü arıyorum. Yardımcı olabilir misin?

Sana 3 adet uygun ürün buldum

Dikdörtgen formunda Benx marka güneş gözlükleri arasından senin için en uygun 3 modeli seçtim. Tüm modeller dikdörtgen çerçeve formuna sahip ve farklı renk seçenekleri sunuyor. Fiyatlar ₺1.399 ile ₺1.800 arasında değişiyor ve hepsi yüksek müşteri memnuniyetine sahip.



BENX SUNGLASSES

Benx Sunglasses Ben.x 9059 C03
52-19 145 Erkek Güneş Gözlüğü

Dikdörtgen Çok Renkli

★★★★★ 5 (2 değerlendirme)

₺1399

İncele



BENX SUNGLASSES

Benx Sunglasses Ben.x Unisex
Güneş Gözlüğü 9055 C01

Dikdörtgen Gri

★★★★★ 5 (1 değerlendirme)

₺1800

İncele



BENX SUNGLASSES

Benx Sunglasses Benx 9059 C06
52 Güneş Gözlüğü

Dikdörtgen Kahverengi

★★★★★ 5 (2 değerlendirme)

₺1410

İncele

Ne arıyorsunuz? Örnek: Dikdörtgen formunda, benx marka güneş gözlüğü arıyorum...

Gönder

Şekil 6.1 : Ürün öneri sistemi arayüz tasarımı çalışması

KAYNAKLAR

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B.** (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M.** (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).
- Aktaş, Ö., Coşkun, B., & Soner, İ.** (2021). Turkish sentiment analysis using machine learning methods: application on online food order site reviews. *Journal of Artificial Intelligence and Data Science, 1*(1), 1-10.
- Ali, H. M. U., Farooq, Q., Imran, A., & El Hindi, K.** (2025). A systematic literature review on sentiment analysis techniques, challenges, and future trends. *Knowledge and Information Systems, 1*-68.
- Artwise AI.** (2023). *modernbert-base-tr-uncased* [Model]. Hugging Face. <https://huggingface.co/artwise-ai/modernbert-base-tr-uncased>
- Aydın, C. R., & Güngör, T.** (2021). Sentiment analysis in Turkish: Supervised, semi-supervised, and unsupervised techniques. *Natural Language Engineering, 27*(4), 455-483
- Ba Alawi, A., & Bozkurt, F.** (2025). Performance analysis of embedding methods for deep learning-based turkish sentiment analysis models. *Arabian Journal for Science and Engineering, 50*(10), 7299-7321.
- Bahdanau, D., Cho, K., & Bengio, Y.** (2015). *Neural machine translation by jointly learning to align and translate*. arXiv:1409.0473.
- Beck, D., Salesky, E., & Wendt, J.** (2021). *W-NUT 2021 Shared Task: Multilingual Lexical Normalization*. In Proceedings of the 7th Workshop on Noisy User-generated Text (W-NUT 2021). Association for Computational Linguistics.
- Behera, G., & Nain, N.** (2022). GSO-CRS: grid search optimization for collaborative recommendation system. *Sādhanā, 47*(3), 158.
- Ben, T. L., Alla, P. C. R., Komala, G., & Mishra, K.** (2023). Detecting sentiment polarities with comparative analysis of machine learning and deep learning algorithms. In *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (pp. 186-190). IEEE.
- Bergstra, J., & Bengio, Y.** (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research, 13*, 281–305.

- Bergstra, J., Yamins, D., & Cox, D.** (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning* (pp. 115–123). PMLR.
- Bischi, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M.** (2023). Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges. *WIREs Data Mining and Knowledge Discovery*, 13(2), e1484.
- Blagec, K., Dorffner, G., Moradi, M., & Samwald, M.** (2020). A critical analysis of metrics used for measuring progress in artificial intelligence. *arXiv preprint arXiv:2008.02577*.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T.** (2017). *Enriching word vectors with subword information*. arXiv:1607.04606.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al.** (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. <https://arxiv.org/abs/2108.07258>.
- Brochu, E., Cora, V. M., & de Freitas, N.** (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.** (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, 1877–1901.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C.** (2017). *New avenues in opinion mining and sentiment analysis*. IEEE Intelligent Systems, 28(2), 15–21.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P.** (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Choi, Y., & Lee, H.** (2017). Data properties and the performance of sentiment classification for electronic commerce applications. *Information Systems Frontiers*, 19(5), 993–1012.
- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Fiedel, N.** (2023). Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240), 1–113.
- Ciftci, B., & Apaydin, M. S.** (2018). A deep learning approach to sentiment analysis in Turkish. In *2018 international conference on artificial intelligence and data processing (IDAP)* (pp. 1–5). IEEE
- Clark, K., Luong, M., Le, Q. V., & Manning, C. D.** (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. *arXiv preprint arXiv:2003.10555*.

- Çakıcı, Ş., Karaduman, D., Çırlan, M. A., & Hürriyetöğlü, A.** (2025). A cross-validation study of Turkish sentiment analysis datasets and tools. *Language Resources and Evaluation*, 1-39
- dbmdz.** (2020a). *electra-base-turkish-mc4-uncased-discriminator* [Model]. Hugging Face. <https://huggingface.co/dbmdz/electra-base-turkish-mc4-uncased-discriminator>
- dbmdz.** (2020b). *bert-base-turkish-128k-uncased* [Model]. Hugging Face. <https://huggingface.co/dbmdz/bert-base-turkish-128k-uncased>
- Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L.** (2023). QLoRA: Efficient Finetuning of Quantized LLMs. *arXiv preprint arXiv:2305.14314*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- Duda, R. O., & Hart, P. E.** (1973). *Pattern classification and scene analysis*. Wiley and Sons, Inc..
- Facebook AI Research.** (2017). *FastText language identification model* [Software documentation]. Retrieved from <https://fasttext.cc/docs/en/language-identification.html>
- Franceschi, L., Donini, M., Perrone, V., Klein, A., Archambeau, C., Seeger, M., Pontil, M., & Frasconi, P.** (2024). Hyperparameter Optimization in Machine Learning. *arXiv preprint arXiv:2410.22854*.
- Gardner, J., Kusner, M., Zhixiang, Weinberger, K., & Cunningham, J.** (2014). Bayesian Optimization with Inequality Constraints. In E. P. Xing & T. Jebara (Eds.), *Proceedings of the 31st International Conference on Machine Learning* (Vol. 32, No. 2, pp. 937–945). PMLR.
- Gemma Team, Riviere, M., Pathak, S., Sessa, P. G., Hardin, C., Bhupatiraju, S., Hussenot, L., Mesnard, T., Shahriari, B., Ramé, A., et al.** (2024b). Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*. <https://arxiv.org/abs/2408.00118>
- Gemma Team.** (2024a). Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*. <https://arxiv.org/abs/2403.08295>.
- George, S.** (2024). Parameter Efficient Fine Tuning (PEFT), inference, and evaluation of LLM model using LoRA. Towards AI. <https://pub.towardsai.net/parameter-efficient-fine-tuning-peft-inference-and-evaluation-of-llm-model-using-lora-03cf9f027c34>
- Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep learning*. MIT Press.
- He, P., Gao, J., & Chen, W.** (2021). DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. *arXiv preprint arXiv:2111.09543*.

- He, P., Liu, X., Gao, J., & Chen, W.** (2020). *DeBERTa: Decoding-enhanced BERT with disentangled attention*. arXiv preprint arXiv:2006.03654.
- Hochreiter, S., & Schmidhuber, J.** (1997). *Long short-term memory*. *Neural Computation*, 9(8), 1735–1780.
- Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S.** (2019). Parameter-Efficient Transfer Learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, 2790–2799.
- Howard, J., & Ruder, S.** (2018). Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W.** (2021). LoRA: Low-Rank Adaptation of Large Language Models. *arXiv preprint arXiv:2106.09685*.
- İncidelen, M., & Aydoğın, M.** (2025). Sentiment Analysis in Turkish Using Language Models: A Comparative Study. *European Journal of Technique (EJT)*, 15(1), 68-74.
- Jain, P. K., Pamula, R., & Srivastava, G.** (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41, 100413. <https://doi.org/10.1016/j.cosrev.2021.100413>
- Jin, H., Huang, M., & Zhu, X.** (2012). Sentiment analysis with multi-source product reviews. In *International Conference on Intelligent Computing* (pp. 301-308). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jurafsky, D., & Martin, J. H.** (2023). *Speech and language processing* (3rd ed., draft). Stanford University.
- Kamali, O.** (2024). *emoji-map* [Dataset]. Hugging Face. <https://huggingface.co/datasets/omarkamali/emoji-map>
- Kanakamedala, V. C., Singh, S. H., & Talasani, R.** (2023). Sentiment Analysis of Online Customer Reviews for Handicraft Product using Machine Learning: A Case of Flipkart. *2023 International Conference on Computer Communication and Informatics (ICCCI)*, 1-6.
- Khan, A., & Baharudin, B.** (2011). Sentiment classification using sentence-level semantic orientation of opinion terms from blogs. In *2011 National Postgraduate Conference* (pp. 1-7). IEEE.
- Kim, Y.** (2014). *Convolutional neural networks for sentence classification*. arXiv:1408.5882.
- Kolchyna, O., Souza, T. T. P., Treleven, P., & Aste, T.** (2015). Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination. *Handbook of Sentiment Analysis in Finance* (pp. 1–32).
- Kumar, R. P., Elakkiya, R., & Rithani, M.** (2023). Automated sentiment classification of amazon product reviews using lstm and bidirectional lstm. In *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)* (pp. 1-6). IEEE.

- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R.** (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *arXiv preprint arXiv:1909.11942*.
- Latha, Y. M., & Rao, B. S.** (2024). Amazon product recommendation system based on a modified convolutional neural network. *ETRI journal*, 46(4), 633-647.
- Law, L.** (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- Liu, B.** (2012). *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies.
- Liu, X., & Wang, C.** (2021). An empirical study on hyperparameter optimization for fine-tuning pre-trained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2286–2300.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V.** (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
- Llama Team, AI@Meta.** (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*. <https://arxiv.org/abs/2407.21783>.
- Lu, Y., Kong, X., Quan, X., Liu, W., & Xu, Y.** (2010). Exploring the sentiment strength of user reviews. In *International Conference on Web-Age Information Management* (pp. 471-482). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Marcondes, F. S., Gala, A., Magalhães, R., Perez de Britto, F., Durães, D., & Novais, P.** (2025). Using ollama. In *Natural Language Analytics with Generative Large-Language Models: A Practical Approach with Ollama and Open-Source LLMs* (pp. 23-35). Cham: Springer Nature Switzerland.
- Masters, D., & Luschi, C.** (2018). Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*. <https://arxiv.org/abs/1804.07612>
- Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., ... & Levenberg, J.** (2018). Mixed precision training. *arXiv preprint arXiv:1710.03740*.
- Microsoft.** (2021). *mdeberta-v3-base* [Model]. Hugging Face. <https://huggingface.co/microsoft/mdeberta-v3-base>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J.** (2013). *Efficient estimation of word representations in vector space*. arXiv:1301.3781.
- Mockus, J.** (1975). On Bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*.

- Mosbach, M., Andriushchenko, M., & Klakow, D.** (2020). On the stability of fine-tuning BERT: Misconceptions, explanations, and strong baselines. *arXiv preprint arXiv:2006.04884*.
- msamilim.** (2025a). *tyroberta-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/tyroberta-turkish-sentiment-optuna-hpo>
- msamilim.** (2025b). *mdeberta-v3-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/mdeberta-v3-turkish-sentiment-optuna-hpo>
- msamilim.** (2025c). *bert-128k-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/bert-128k-turkish-sentiment-optuna-hpo>
- msamilim.** (2025d). *turkishbertweet-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/turkishbertweet-turkish-sentiment-optuna-hpo>
- msamilim.** (2025e). *modernbert-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/modernbert-turkish-sentiment-optuna-hpo>
- msamilim.** (2025f). *electra-turkish-sentiment-optuna-hpo* [Model]. Hugging Face. <https://huggingface.co/msamilim/electra-turkish-sentiment-optuna-hpo>
- Mullen, T., & Collier, N.** (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 412-418).
- Najafi, A., & Varol, O.** (2023). TurkishBERTweet: Fast and Reliable Large Language Model for Social Media Analysis. *arXiv preprint arXiv:2311.18063*.
- Narayanan, V., Arora, I., & Bhatia, A.** (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *Intelligent Data Engineering and Automated Learning IDEAL 2013 Lecture Notes in Computer Science Volume 8206*, 194-201. *arXiv:1305.6143*.
- Ng, A. Y.** (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the Twenty-First International Conference on Machine Learning*.
- Nguyen, D. Q., Vu, T., & Nguyen, A. T.** (2020). BERTweet: A pre-trained language model for English Tweets. *arXiv preprint arXiv:2005.10200*.
- Ollama.** (2025). *Library*. <https://ollama.com/library>
- Özdemir, A. O., Giritli, E. B., & Can, Y. S.** (2024). Sentiment Analysis for Hotel Reviews in Turkish by Using LLMs. In *2024 9th International Conference on Computer Science and Engineering (UBMK)* (pp. 210-214). IEEE
- Pang, B., Lee, L., & Vaithyanathan, S.** (2002). *Thumbs up? Sentiment classification using machine learning techniques*. Proceedings of EMNLP.

- Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A.** (2024). The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities. *arXiv preprint arXiv:2408.13296*.
- Passos, D., & Mishra, P.** (2022). A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems*, 223, 104520.
- Pedregosa, F.** (2016). Hyperparameter optimization with approximate gradient. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of The 33rd International Conference on Machine Learning* (Vol. 48, pp. 737–746). PMLR.
- Pennington, J., Socher, R., & Manning, C. D.** (2014). *GloVe: Global vectors for word representation*. arXiv:1406.0728.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L.** (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*. (NAACL 2018).
- Powers, D. M.** (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I.** (2018). *Improving language understanding by generative pre-training*.
- Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R.** (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. *Artificial Intelligence Laboratory; Massachusetts Institute of Technology; Cambridge, MA*.
- Rogers, A., Kovaleva, O., & Rumshisky, A.** (2020). A Primer in BERTology: What we know about how BERT works. *arXiv preprint arXiv:2002.12327*.
- Sak, H., Güngör, T., & Saraçlar, M.** (2007). Morphological disambiguation of Turkish text with perceptron algorithm. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 107-118). Berlin, Heidelberg: Springer Berlin Heidelberg
- Santur, Y.** (2019). Sentiment analysis based on gated recurrent unit. In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-5). IEEE.
- Schweter, S.** (2020). BERTurk - BERT models for Turkish. Zenodo. doi:10.5281/zenodo.3770924.
- Shekhar, S., Bansode, A., & Salim, A.** (2022). A Comparative study of Hyper-Parameter Optimization Tools. *arXiv preprint arXiv:2201.06433*..
- Singh, C., Imam, T., Wibowo, S., & Grandhi, S.** (2022). A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews. *Applied Sciences*, 12(8), 3709.
- Snoek, J., Larochelle, H., & Adams, R. P.** (2012). Practical Bayesian Optimization of Machine Learning Algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 2951–2959). Curran Associates, Inc..

- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M.** (2011). *Lexicon-based methods for sentiment analysis*. *Computational Linguistics*, 37(2), 267–307.
- Tan, K. L., Lee, C. P., & Lim, K. M.** (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7), 4550.
- Teke, B., Zamir, G., Budak, A. B., & Aksakallı, I. K.** (2025). BERTurk-Based Sentiment Analysis on E-Commerce Multi Domain Product Reviews. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 25(3), 497-509.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G.** (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*. <https://arxiv.org/abs/2302.13971>.
- Trendyol.** (2022). *ty Roberta* [Model]. Hugging Face. <https://huggingface.co/Trendyol/ty Roberta>
- ÚFAL.** (2021). *W-NUT 2021: Multilingual Lexical Normalization (MultiLexNorm 2021) – Improving multilingual lexical normalization by fine-tuning ByT5* [GitHub repository]. Retrieved from <https://github.com/ufal/multilexnorm2021>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.** (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- Vinodhini, G., & Chandrasekaran, R. M.** (2014). Measuring the quality of hybrid opinion mining model for e-commerce application. *Measurement*, 55, 101-109.
- Vinodhini, G., & Chandrasekaran, R. M.** (2016). A sampling based sentiment mining approach for e-commerce applications. *Information Processing & Management*, 53(1), 223-236
- VRLLab.** (2023). *TurkishBERTweet* [Model]. Hugging Face. <https://huggingface.co/VRLLab/TurkishBERTweet>
- Vural, A. G., Cambazoglu, B. B., Senkul, P., & Tokgoz, Z. O.** (2013). A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. In *Computer and information sciences III: 27th international symposium on computer and information sciences* (pp. 437-445). London: Springer London.
- Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Cooper, N., Adams, G., Howard, J., & Poli, I.** (2024). *Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference*. *arXiv preprint arXiv:2412.13663*.
- Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V.** (2022). Finetuned Language Models are Zero-Shot

Learners. In *The Tenth International Conference on Learning Representations, ICLR 2022*.

- Wilson, T., Wiebe, J., & Hoffmann, P.** (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 347–354). Vancouver, British Columbia, Canada: Association for Computational Linguistics.
- Xia, R., Xu, F., Zong, C., Li, Q., Qi, Y., & Li, T.** (2015). Dual sentiment analysis: Considering two sides of one review. *IEEE transactions on knowledge and data engineering*, 27(8), 2120-2133.
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V.** (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. *arXiv preprint arXiv:1906.08237*.
- Young, T., Hazarika, D., Poria, S., & Cambria, E.** (2018). *Recent trends in deep learning based natural language processing*. IEEE Computational Intelligence Magazine, 13(3), 55–75.
- Zümbereöglu, K. B., Dik, S. Z., Karadeniz, B. S., & Sahmoud, S.** (2025). Towards Better Sentiment Analysis in the Turkish Language: Dataset Improvements and Model Innovations. *Applied Sciences* (2076-3417), 15(4).



ÖZGEÇMİŞ

Ad-Soyad : Muhammet Şamil İKİZOĞLU

ÖĞRENİM DURUMU:

- **Lisans** : 2018, İstanbul Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Elektronik ve Haberleşme Mühendisliği Programı
- **Yüksek Lisans** : 2026, İstanbul Teknik Üniversitesi, İşletme Mühendisliği Fakültesi, İşletme Mühendisliği Programı