TÜRKÇE METİNLERDE İRONİ VE İĞNELEME İÇEREN İFADELERİN DÖNÜŞTÜRÜCÜ TEMELLİ KOLEKTİF ÖĞRENME YÖNTEMLERİ İLE TESPİT EDİLMESİ

Murat ESER

T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE METİNLERDE İRONİ VE İĞNELEME İÇEREN İFADELERİN DÖNÜŞTÜRÜCÜ TEMELLİ KOLEKTİF ÖĞRENME YÖNTEMLERİ İLE TESPİT EDİLMESİ

Murat ESER
0000-0001-8052-6587

Doç. Dr. Metin BİLGİN
(Danışman)
0000-0002-4216-0542

DOKTORA TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

BURSA – 2026
Her Hakkı Saklıdır

ÖZET

Doktora Tezi

TÜRKÇE METİNLERDE İRONİ VE İĞNELEME İÇEREN İFADELERİN DÖNÜŞTÜRÜCÜ TEMELLİ KOLEKTİF ÖĞRENME YÖNTEMLERİ İLE TESPİT EDİLMESİ

Murat ESER

Bursa Uludağ Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Metin BİLGİN

İroni ve iğneleme, söylenen ile söylenmek istenen ifadelerin arasındaki uyumsuzluğu vurgulayan ifade biçimleridir. Bu ifadelerin tespit edilmesi, sosyal medya gibi kullanıcı merkezli platformlarda önemli bir metin madenciliği problemidir. Çevrimiçi olarak büyük miktarda metinsel veri bu kullanıcıların duygu eğilimlerine ilişkin ipuçları taşımaktadır. Bu nedenle ironik ve iğneleyici ifadeleri belirlemek kişilerin duygularını daha iyi tanımlamak için önemlidir. Ancak ironik ve iğneleyici ifadelerin aynı anda tespit edilmeye çalışılması, birbirlerine yakın anlamlar taşıması nedeniyle oldukça zordur. Bu ifadelerin kullanımının yaygınlaşması nedeniyle de doğal dil işleme alanında önemli bir araştırma konusu haline gelmiştir.

İronik ve iğneleyici ifadelerin tespit edilmesi amacıyla bu çalışmada, Türkçe metinler üzerinde BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk ve RoBERTaTurk olmak üzere beş farklı önceden eğitilmiş dönüştürücü tabanlı modeller ve kolektif öğrenme yöntemi, metinsel olarak aynı ancak bağlam uzunluğu açısından farklı bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümeleri kullanılarak performansları karşılaştırılmaktadır. Performansın artırılması amacıyla, yapay tavşan optimizasyon algoritmasına dayanan kolektif öğrenme yöntemiyle model çıktıları birleştirilerek nihai tahminler elde edilmiştir. Elde edilen deneysel sonuçlar, veri kümelerindeki bağlam genişledikçe modellerin daha iyi tahminler yaparak performans metriklerinde iyileştirici bir etki yarattığını göstermektedir. BERTurk ve ConvBERTurk modelleri tüm veri kümelerinde diğer modellere göre daha iyi performans göstermiştir. Kolektif öğrenme yöntemi, bağlam genişledikçe performansı artmış tüm metriklerde bireysel modellerin üzerinde performans göstererek, tüm-gram veri kümesinde %76,71 doğruluk, %74,64 kesinlik, %73,29 duyarlılık ve %73,96 F1-skoru değerlerini elde etmiştir. Bu çalışma Türkçe gibi düşük kaynaklı dillerde kolektif öğrenmenin ve bağlamın ironi ve iğnelemenin tespitinde belirleyici olduğunu ortaya koymaktadır.

Anahtar Kelimeler: İroni, iğneleme, metin sınıflandırma, kolektif öğrenme, dönüştürücü

2026, vii + 102 sayfa.

ABSTRACT

PhD Thesis

IDENTIFICATION OF EXPRESSIONS CONTAINING IRONY AND SARCASM IN TURKISH TEXTS WITH TRANSFORMER-BASED ENSEMBLE LEARNING METHODS

Murat ESER

Bursa Uludağ University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Metin BİLGİN

Irony and sarcasm are expressive forms that highlight the discrepancy between what is literally said and what is actually intended. Accurately identifying such expressions constitutes a significant text mining challenge, particularly on user-centric platforms such as social media. Online textual data, generated in large volumes by users, often contains valuable cues about their emotional tendencies. Consequently, the detection of ironic and sarcastic expressions is crucial for a more precise understanding of individuals' sentiments. However, attempting to detect ironic and sarcastic expressions simultaneously is particularly challenging due to their close semantic similarity. The growing prevalence of these expressions has established irony and sarcasm detection as a prominent research topic in the field of natural language processing.

In this study, five different pre-trained transformer-based models—namely BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk, and RoBERTaTurk—along with an ensemble learning approach, are compared in terms of their performance on Turkish texts. The evaluation utilizes textual datasets that share identical content but differ in contextual scope: one-gram, two-gram, three-gram, four-gram, and all-gram datasets. To enhance performance, an ensemble learning method based on the Artificial Rabbit Optimization algorithm is employed to combine the model outputs and derive final predictions. The experimental results demonstrate that as the context expands across the datasets, the models produce increasingly accurate predictions, leading to consistent improvements in performance metrics. Among the individual models, BERTurk and ConvBERTurk consistently outperform the others across all datasets. The ensemble learning approach surpasses the individual models in every metric, with performance gains becoming more pronounced as context length increases. On the all-gram dataset, it achieves 76.71% accuracy, 74.64% precision, 73.29% recall, and an F1-score of 73.96%. This study highlights the decisive role of ensemble learning and contextual scope in the detection of irony and sarcasm, particularly in low-resource languages such as Turkish.

Key words: Irony, sarcasm, text classification, ensemble learning, transformer

2026, vii + 102 pages.

TEŞEKKÜR

Doktora eğitimim süresince çalışmalarımda önderlik eden, akademik, kültürel ve sosyal anlamda değerli katkılarıyla yön veren, çalışmalarım sırasında beni sürekli cesaretlendiren ve desteğini her zaman hissettiğim danışman hocam Sayın Doç. Dr. Metin BİLGİN'e, tez çalışmalarım boyunca fikirleriyle ve tecrübeleriyle desteklerini esirgemeyen tez izleme komisyonu üyeleri Sayın Prof. Dr. Fatih Mehmet AMASYALI ve Sayın Prof. Dr. Fahri VATANSEVER hocalarıma teşekkürlerimi sunarım.

Bu zorlu süreçte bana her zaman destek veren ve yanımda olan değerli eşime, sevgili kızıma, beni büyüten, yetiştiren, eğitim hayatım boyunca her zaman destek olan annem, babam ve kardeşime teşekkürlerimi sunarım.

Murat ESER
29/01/2026

İÇİNDEKİLER

Sayfa

ÖZET i
ABSTRACT ii
TEŞEKKÜR iii
İÇİNDEKİLER iv
SİMGELER ve KISALTMALAR DİZİNİ v
ŞEKİLLER DİZİNİ vi
ÇİZELGELER DİZİNİ vii
1. GİRİŞ 1
1.1. Tezin Amacı 3
1.2. Araştırmanın Katkıları 4
1.3. Araştırma Soruları 4
1.4. Tezin Organizasyonu 5
2. KAYNAK ARAŞTIRMASI 7
3. MATERYAL ve YÖNTEM 29
3.1. Veri Kümesi 30
3.1.1. Veri kümesinin oluşturulması 31
3.1.2. Veri kümesinin etiketlenmesi 32
3.1.3. Alt veri kümelerinin oluşturulması 35
3.1.4. Eğitim ve test veri kümelerinin hazırlanması 37
3.2. Modelleme 38
3.2.1. Destek vektör makineleri 40
3.2.2. BERTurk 40
3.2.3. ConvBERTurk 43
3.2.4. DistilBERTurk 44
3.2.5. ELECTRATurk 45
3.2.6. RoBERTaTurk 46
3.2.7. Yapay tavşan optimizasyon algoritması 46
3.2.8. Kolektif öğrenme 50
3.3. Önerilen Yöntem 53
3.4. İroni ve İğneleme Tespitinde Kullanılan Performans Metrikleri 55
4. BULGULAR 57
4.1. Modellerin Bireysel Performans Değerlendirmesi 57
4.2. Topluluk Öğrenmesi Performans Değerlendirmesi 72
5. TARTIŞMA ve SONUÇ 84
KAYNAKLAR 87
ÖZGEÇMİŞ 102

SİMGELER ve KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
BERT	Bidirectional Encoder Representations from Transformers (Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri)
BiGRU	Bidirectional Gated Recurrent Unit (Çift Yönlü Kapılı Yinelemeli Birim)
CNN	Convolutional Neural Network (Evrişimsel Sinir Ağı)
DistilBERT	Distilled BERT (Damıtılmış BERT)
GRU	Gated Recurrent Unit (Kapılı Yinelemeli Birim)
KNN	K-nearest Neighbors K En Yakın Komşu Algoritması)
RoBERTa	Robustly Optimized BERT Approach (Dayanıklı Biçimde Optimize Edilmiş BERT Yaklaşımı)
SVM	Support Vector Machines (Destek Vektör Makineleri)
GNN	Graph Neural Networks (Graf Sinir Ağları)
DDİ	Doğal Dil İşleme
RNN	Recurrent Neural Network (Yinelemeli Sinir Ağı)
BiRNN	Bidirectional Recurrent Neural Networks (Çift Yönlü Yinelemeli Sinir Ağı)
LSTM	Long Short Term Memory (Uzun Kısa Süreli Bellek)
BiLSTM	Bidirectional Long Short Term Memory (Çift Yönlü Uzun Kısa Süreli Bellek)
TF-IDF	Term Frequency-Inverse Document Frequency (Terim Frekansı–Ters Belge Frekansı)
TF	Term Frequency (Terim Frekansı)
XLNET	Extra Long Network (Ekstra Uzun Ağ)
LLM	Large Language Model (Büyük Dil Modeli)
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately (Belirteç Değiştirmelerini Yüksek Doğrulukla Sınıflandıran Bir Kodlayıcının Etkin Öğrenimi)
ELMo	Embeddings from Language Models (Dil Modellerinden Elde Edilen Gömme Temsilleri)
Word2vec	Word to Vector (Kelime Vektörleri)
GloVe	Global Vectors for Word Representation (Kelime Gösterimi için Küresel Vektörler)
MDMSC	Maskeli Dil Modellemesi ve Sonraki Cümle Çıkarımı
AVT	Adlandırılmış Varlık Tanıma
SSCVS	Stanford Soru-Cevap Veri Kümesi
MDM	Maskeli Dil Modelleme
SCT	Sonraki Cümle Tahmini
LKonv	Lokal Evrişim
Konv	Evrişim
CLS	Sınıflandırma Birimi
SEP	Ayırıcı

ŞEKİLLER DİZİNİ

			Sayfa
Şekil	3.1.	Çalışmada izlenilen süreç ve çalışmanın akış şeması…….	29
Şekil	3.2.	Veri kümesi oluşturma işleminin akış şeması…….	31
Şekil	3.3.	Sınıflara göre temel istatistikler……..	34
Şekil	3.4.	Veri kümesinde kullanılan metinlerin kategorileri…….	34
Şekil	3.5.	Dönüştürücü Mimarisi (Çağla, 2024)…….	41
Şekil	3.6.	BERT modelinin ön eğitim ve ince ayar (Devlin vd., 2019'dan değiştirilerek alınmıştır)..…….....	43
Şekil	3.7.	ConvBERT mimari diyagramı (Jiang vd., 2020'den değiştirilerek alınmıştır)……..	44
Şekil	3.8.	DistilBERT modelinin mimarisi (J. Li vd., 2024)……..	45
Şekil	3.9.	ELECTRA modelinin token yer değiştirmesi (Çağla, 2024)…..	45
Şekil	3.10.	ARO algoritması ile oluşturulan topluluk modelinin akış şeması (L. Wang vd., 2022'den değiştirilerek alınmıştır)……..	47
Şekil	3.11.	Tavşanların hareketleri A) 2-D B) 3-D……...	48
Şekil	3.12.	Enerji faktörü A'ya dayalı arama mekanizması (L. Wang vd., 2022'den değiştirilerek alınmıştır)……..	49
Şekil	3.13.	Kolektif öğrenme gösterimi……	52
Şekil	3.14.	Önerilen kolektif öğrenme yöntemi……	54
Şekil	3.15.	Çok sınıflı karmaşıklık matrisi……	55
Şekil	4.1.	Modellerin performans gelişim eğrisi…….	72
Şekil	4.2.	Kolektif öğrenme bir-gram veri kümesi karmaşıklık matrisi…..	74
Şekil	4.3.	Kolektif öğrenme iki-gram veri kümesi karmaşıklık matrisi….	75
Şekil	4.4.	Kolektif öğrenme üç-gram veri kümesi karmaşıklık matrisi…..	77
Şekil	4.5.	Kolektif öğrenme dört-gram veri kümesi karmaşıklık matrisi…	78
Şekil	4.6.	Kolektif öğrenme tüm-gram veri kümesi karmaşıklık matrisi…….	80

ÇİZELGELER DİZİNİ

			Sayfa
Çizelge	2.1.	İğneleme ve ironi tespiti için gerçekleştirilen çalışmalarda kullanılan yöntemler, sonuçlar ve kullanılan veri kümeleri……..	26
Çizelge	3.1.	Sınıflara göre veri kümesi mimarileri……	36
Çizelge	3.2.	Yapılan çalışmalarda kullanılan yöntemlerin kategorilere göre sınıflandırılması…….	39
Çizelge	4.1.	Modellerin bir-gram veri kümesinde bireysel performans sonuçları…….	58
Çizelge	4.2.	Modellerin bir-gram veri kümesinde sınıflara göre performans sonuçları…….	59
Çizelge	4.3.	Modellerin iki-gram veri kümesinde bireysel performans sonuçları.....	60
Çizelge	4.4.	Modellerin iki-gram veri kümesinde sınıflara göre performans sonuçları…….	61
Çizelge	4.5.	Modellerin üç-gram veri kümesinde bireysel performans sonuçları…….	62
Çizelge	4.6.	Modellerin üç-gram veri kümesinde sınıflara göre performans sonuçları…….	63
Çizelge	4.7.	Modellerin dört-gram veri kümesinde bireysel performans sonuçları…….	64
Çizelge	4.8.	Modellerin dört-gram veri kümesinde sınıflara göre performans sonuçları……	65
Çizelge	4.9.	Modellerin tüm-gram veri kümesinde bireysel performans sonuçları.	66
Çizelge	4.10.	Modellerin tüm-gram veri kümesinde sınıflara göre performans sonuçları…….	67
Çizelge	4.11.	Veri kümelerine göre en iyi performans sonuçları…….	68
Çizelge	4.12.	Veri kümelerinde en başarılı modellerin performans sonuçları….	69
Çizelge	4.13	Veri kümelerinde düşük performanslı modellerin sonuçları…….	71
Çizelge	4.14.	Modellerin bir-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları……...	73
Çizelge	4.15.	Modellerin iki-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları……...	75
Çizelge	4.16.	Modellerin üç-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları……...	76
Çizelge	4.17.	Modellerin dört-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları……..	78
Çizelge	4.18.	Modellerin tüm-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları……..	79
Çizelge	4.19.	Kolektif öğrenme yönteminin veri kümesi türüne göre performans sonuçları…….	81
Çizelge	4.20.	Optimizasyon algoritmasının ürettiği ağırlıkların istatistikleri….	82

Çizelge 4.21. Veri kümeleri üzerinde SarcaTurkGPT'nin elde ettiği performans metrikleri……. 83

1. GİRİŞ

İnsanoğlu, dünya genelinde her saniye veri akışının olduğu ve bu verilerin el ile yapılan yöntemlerle çok azının anlaşılabildiği bir çağda yaşamaktadır (Junaid vd., 2022). İnternetin ortaya çıkışıyla birlikte büyük miktarda metinsel veri kullanılabilir hale gelmiştir (Bharti vd., 2016; Reyes vd., 2012). Bunların içerisinde sayısal veriler belirli istatistiksel araçlarla görselleştirilerek analiz edilebilmektedir ancak aynı durum metinler söz konusu olduğunda pek mümkün görünmemektedir (Junaid vd., 2022). Teknoloji geliştikçe, insanlara basit metin mesajları ve mesaj panolarından resimler ve videolar gibi daha ilgi çekici ve sürükleyici kanallara kadar giderek daha fazla etkileşim imkânı ortaya çıkmıştır (Bharti vd., 2016). Web siteleri, haberler, bloglar, kitaplar, bilimsel makaleler ve sosyal medya gibi kaynaklarda var olan bilgi yığınını özümsemek giderek daha zor hale gelmektedir (Baykara ve Güngör, 2023). İnsan kapasitesinin büyük hacimli metin verilerini işleme konusunda yetersiz kalmasından dolayı, Doğal Dil İşleme (Natural Language Processing-DDİ) bu tür verileri daha anlayışlı ve hızlı bir şekilde işlenebilir hale getirmek için insan zekâsında olduğu gibi ancak daha hızlı bir biçimde ve insan zekâsının bir anda algılayabileceğinden çok daha büyük hacimde okumaya, yorumlamaya ve anlamaya yönelik kullanılmaktadır (Junaid vd., 2022). Bu nedenle, son birkaç on yılda, ilgili bilgi parçalarını elde etme sürecini basitleştirmeyi ve daha verimli hale getirmeyi amaçlayan metin özetleme, popüler bir araştırma alanı olarak ortaya çıkmıştır (Baykara ve Güngör, 2023).

DDİ, cümle sınıflandırma, cümle çifti sınıflandırma, soru sorma ve cevaplama, adlandırılmış varlık ilişki sınıflandırma başta olmak üzere birçok DDİ probleminin üstesinden gelmek için kullanılmaktadır (Junaid vd., 2022). Karmaşık bir süreç olan sözlü (H. M. Kumar ve Harish, 2018; Reyes ve Rosso, 2012) ve yazılı (Fersini vd., 2014; G. Wang vd., 2014) iletişim, ortak bir yaşantı ve geçmiş paylaşmanın yanı sıra bağlamsal anlamın ötesinde bilgi çıkarabilmeyi gerektirir. Pek çok iletişimsel eylem, anlamın tamamını çözebilmek için dil bilimsel olarak ifade edilmeyen bilgileri ima eder: eğer dinleyici bu bilgileri çıkaramıyorsa, iletişimsel süreç tamamlanmamış demektir (Reyes ve Rosso, 2012). Bu metinlerin içerisinde söz sanatlarının yoğun olduğu içerikler bulunmaktadır (Davidov vd., 2010; Filatova, 2012; Poria vd., 2016; Sulis vd., 2016). Söz sanatları; ironi, iğneleme, hiciv, parodi, vb. kavramları

içermektedir (Ravi ve Ravi, 2018). Ancak söz sanatlarının insanlar tarafından anlaşılması otomatik metin ve veri madenciliği sistemleri kullanılarak bile zor bir husustur (Ravi ve Ravi, 2017), çünkü ironi ve iğneleme gibi kinayeli ifadelerin çarpıcı özelliği, gerçek anlamı ile kastedilen anlamı arasındaki boşluğu kapatmayı ve köprü kurmayı zorlaştırmasıdır. İroni ve iğneleme tespiti için bu tür ifadelerin bağlamındaki temel sorun söylenen ile söylenmek istenen arasındaki çelişkiyi tespit etmek gerekmektedir (Ashwitha vd., 2020). İroni ve iğneleme sosyal medyada, çevrimiçi forumlarda ve internet sitelerinde yaygın olduğundan, çevrimiçi metinleri analiz etmeye çalışan modellerin bunu tespit edebilmesi ve anlayabilmesi kritik önem taşımaktadır. İğneleyici ifadeler konuşmacının gerçek niyetinden farklı bir anlam taşıyan, bir kişinin duygularını incitmek veya mizahi şekilde eleştirmek amacıyla söylediklerinin açıkça tersini ifade eden sözlerin kullanılması olarak tanımlanmaktadır. Üstü kapalı olarak onur kırıcı, üzüntü verici söz söylemek iğnelemenin temel amacıdır. İronik ifadeler bir durumun gerçekte hissedilen duygudan farklı söylenmesidir. Bir beklenti olduğunda tam tersi bir durumun gerçekleşmesi olarak da ifade edilebilir. Kavramsal açıdan ironi, içerisinde alayın belirgin olduğu ve farklı anlamları da bünyesinde barındıran kinayeli ve dolaylı bir anlatım tekniğidir. İroninin en yaygın kullanım biçimi, kişinin gayri ciddi bir şeyi ciddiymiş gibi ifade etmesidir (Kader vd., 2022). İroni ve iğneleme insan dilinin öne çıkan özelliklerindendir, incelikli ifade ve sosyal etkileşimde derin kökleri vardır. Her iki olgu da kastedilen anlamla çelişen bir şey söylemeyi içerir, ancak farklı şekilde kullanılırlar ve farklı çıkarımlar taşırlar. Bağlam hem ironi hem de iğnelemenin algılanması ve anlaşılmasında vazgeçilmez bir rol oynamaktadır. Rohanian ve arkadaşları, beklentilere uymamayı ve ardından duygusal tepkiler ortaya çıkarmayı içeren durumsal ironinin karmaşık doğasına işaret etmektedir (Rohanian vd., 2018). İğneleme ve ironi bazen birbirinin yerine kullanılabileceği düşünülmektedir. Ancak Sulis ve arkadaşları, iğnelemenin genellikle daha açık bir şekilde agresif ve keskin bir tonda olduğunu, ironinin ise daha ince, daha alaycı bir nitelik taşıyabileceğini belirterek bunları birbirinden ayırmaktadır (Sulis vd., 2016). Başkalarının duygularını incitmek veya bir şeye mizahi bir şekilde saldırmak için yapılan ve söylenenin tersi anlamına gelen yorumların kullanımı; bu tanımla iğnelemenin günlük hayatımızda çok önemlidir (Chauhan vd., 2022). İroni, basit bir tanımla dilin yaratıcı kullanımıdır (Taşlıoğlu, 2014) ve genellikle küçümseme için kullanılan, konuşmacının gerçek anlamının tersi bir mesaj

ilettiği, genellikle belirli bir hedefle alay etme niyetinde olduğu bir iletişim şeklidir (Yue vd., 2023).

İroni ve iğneleme gibi ifadelerin kullanım biçimi yüz ifadelerinde, jestlerde ve metinlerde görülmektedir. Bu ifadeler özneldir; duruma, kişilere, dile ve kültüre göre değişir ve çoğunlukla olumlu ile olumsuz unsurları birlikte içerir (Ren vd., 2020). Bu durum, bağlamın doğru biçimde anlaşılması ve bağlama özgü duyguların tanımlanmasını gerekli kılmaktadır (Ashwitha vd., 2020). Bu yönüyle bu tür analizler, son zamanlarda bilgisayar bilimcilerinin ilgisini çekmektedir (Taşlıoğlu, 2014). İğneleme, kesici bir tonla kullanılan saldırgan bir ironi türüdür (Attardo, 2000). Dünya genelinde, iğnelemenin günlük yaşama dahil edilmesinde gözle görülür bir artış eğilimi vardır. Bu eğilim, iğnelemenin günlük yaşamda sıkça kullanılmasına, ancak daha spesifik olarak sosyal medya ve internete bağlanabilir (Goel vd., 2022). Siyasette kutuplaşmanın artması (Enyedi, 2016) ve artan popülizm (Moffitt, 2016), siyasette duyguların analizine rolüne yönelik bilimsel ilgiyi artırmıştır. Bu bilimsel eğilim; kamu politikaları üzerine tartışmalar (Lai vd., 2015), meclis konuşmalarındaki kızgınlık, korku, mutluluk, üzüntü vb. duygu analizleri (Eskişar ve Çöltekin, 2022), meclis tartışmalarının analizi (Elçi, 2019), politik tartışmalar (Stranisci vd., 2016) ve siyasi liderlerin konuşmalarının içerik analizi (Devran, 2016) gibi araştırmalarda kendini göstermektedir. Kinayeli iletişimi içeren bir şemsiye terim olarak ironi ve iğneleme kavramları kullanılmaktadır. Literatürde iğnelemenin, ironiye göre daha agresif olma eğiliminde olduğuna dair bilimsel kanıtlar (Chaudhari ve Chandankhede, 2018; Clift, 1999; Lee ve Katz, 1998) olması nedeniyle daha sert ve kırıcı ifadeler iğneleme olarak değerlendirilirken, söylenilenle söylenmek istenilenin birbirinin zıttı olduğu ifadeler (Freitas vd., 2014) ironi olarak değerlendirilmiştir.

1.1. Tezin Amacı

Tez çalışması doğrultusunda ironik ve iğneleyici ifadeler üzerine gerçekleştirilen literatür araştırması, kinayeli iletişim içeren bu ifadelerin birlikte tespitinin çok az olduğu, yapılan çalışmaların genellikle kaynak açısından zengin diller üzerine yoğunlaştığını göstermektedir. Türkçe diline ait kapsamlı bir çalışma veri kümesi ile çalışma olmadığı yapılan literatür araştırması sonucunda ortaya çıkmaktadır. Yetersiz

veri kaynağına sahip yerel dilde yapılan çalışmalar dile ait kültürel bağlamı daha iyi ele almaktadır.

Türkçe metinlerde ironi ve iğneleme içeren ifadelerin aynı anda tespiti çalışmalarının sınırlı sayıda olması, bu ifadelerin tespitine yönelik hazırlanmış yeterli ve kapsamlı veri kümesinin bulunmaması, olumlu gibi görünen ancak olumsuz anlamlar taşıyabilecek kinayeli ifadelerin tespitinin yazılı iletişimde önemli bir yere sahip olması, duygu analizi çalışmalarında başarıyı artırabilecek olması tez çalışmasının temelini oluşturmaktadır. Çalışma ile köşe yazılarının metin işleme teknikleri ile incelenmesi ve analiz edilmesi sonucunda ironik ve iğneleyici ifade barındırma durumlarının tespit edilmesi amaçlanmaktadır. Türkçe ironik ve iğneleyici ifadelerin tespiti için doğruluk düzeyi yüksek ve güvenilir bir yapay zekâ modeli ortaya konulması amaçlanmıştır.

1.2. Araştırmanın Katkıları

Tez çalışmasında Türkçe köşe yazılarında bağlam temelli ironik ve iğneleyici ifadelerin aynı anda tespitine yönelik literatürdeki boşluğu doldurması amaçlanmaktadır. Türkçe gibi az kaynaklı dillerde doğal dil işleme çalışmalarının ve kapsamlı veri kümelerinin sınırlı olduğu görülmüştür. Türkçe dili üzerinde doğal dil işleme çalışmalarında bulunan bu araştırma boşluğunun doldurulması amacıyla:

- Köşe yazılarından Türkçe diline ait dengeli ve 3 sınıflı veri kümesi oluşturulması
- İroni ve iğnelemenin aynı anda analizi
- Bağlam duyarlı modellerin geliştirilmesi
- Gelecekte yapılacak duygu analizi çalışmalarına katkı vermesi

beklenmektedir.

1.3. Araştırma Soruları

Köşe yazılarında bulunan ironik ve iğneleyici ifadeler, dilin kullanımına göre anlaşılması zordur. Ayrıca biçimsel, anlamsal, bağlamsal ve söz dizimsel gibi her dile ait kavramları, dile ait kültürel geçmiş gibi dilin anlaşılması ve yorumlanmasına önemli

etki edebilecek durumlar da bu ifadelerin tespitinde önemli rol oynamaktadır. Bu ifadelerin Türkçe yazılı metinlerinde tespitinin yapılabilmesi amacıyla:

- Bağlam temelli veri kümelerinin performans üzerine etkisi var mı?
- Türkçe ön eğitimli dönüştürücü temelli dil modellerinin performans çıktıları nedir?
- Klasik makine öğrenmesi yöntemleri ve kelime gömme yöntemleri dil modellerine kıyasla model boyutu açısından daha faydalı mı?
- Kolektif öğrenme yönteminin performans üzerindeki etkisi nedir?
- Farklı bağlam uzunluklarına göre kolektif öğrenme yöntemi temel modellere göre performans artışı sağlar mı?

soruları doğrultusunda çalışmalar gerçekleştirilmiştir.

1.4. Tezin Organizasyonu

Bu tez çalışması 5 bölüm olarak yapılandırılmıştır. Birinci bölümde ironi ve iğneleyici ifadeler, birbirlerine benzerlikleri ve farklılıkları, Türkçe dilindeki yeri ve neden önemli olabileceğinden bahsedilmiştir. İlk bölümde temel bilgilere ek olarak ortaya konulan problem, çalışmanın amacı, araştırmanın katkıları ve çalışma sürecinde hedefe yönelik araştırma soruları açıklanmıştır. İkinci bölümde sosyal medya üzerinden oluşturulan veri kümeleri ile farklı yöntemler kullanılarak yapılan çalışmalar incelenmiştir. Üçüncü bölümde çalışma kapsamında oluşturulan veri kümelerinin elde edilmesi, etiketlenmesi, bağlamsal aşamalı veri kümelerinin hazırlanması, veri ön işlem aşamaları ve modellerin eğitiminde kullanılacak olan verilerin hazırlanma yöntemleri sunulmuştur. Çalışmada kullanılan Türkçe ön eğitimli dil modellerinin belirlenmesi bu bölümde gerçekleştirilmiş olup tanımlamaları yapılmıştır. Kullanılan algoritmaların ve dil modellerinin analizi için seçilen performans ölçüm metrikleri yine bu bölümde açıklanmıştır. Çalışmada kolektif öğrenme olarak kullanılan optimizasyon algoritması anlatılmış ve sınıflandırma problemlerinde birleştirici yöntem olarak kullanımı bu bölümde açıklanmıştır. Dördüncü bölümde hazırlanan veri kümeleri üzerinde Türkçe ile önceden eğitilmiş dil modelleri ve klasik makine öğrenmesi olarak Destek Vektör Makineleri (Support Vector Machines-SVM) yöntemlerinin performans analizi için gerçekleştirilen deneysel çalışmalar hakkında detaylı analiz, değerlendirmeler ve

karşılaştırmalı sonuçlar yapılmıştır. Kolektif öğrenmenin analiz sonuçlarına etkisi incelenmiştir. Son bölümde, çalışmanın sonuçları özetlenmiş, gelecekte yapılabilecek çalışmalara değinilmiş ve önerilerde bulunulmuştur.

2. KAYNAK ARAŞTIRMASI

Makine öğrenimi modelleri doğal dil işleme uygulamaları için giderek daha popüler hale geldikçe, bu modellerin çevrimiçi metinlerdeki ironi ve iğneleyici ifadeleri tanıyabilmesine duyulan ihtiyaç da artmaktadır (Lad, 2023). Tam bu noktada mevcut haliyle kısıtlılıkları olsa da (Ravi ve Ravi, 2017), DDİ, bu probleme yönelik ciddi çözümlerin bulunduğu bir alan olarak ortaya çıkmaktadır (Junaid vd., 2022). Bu uygulamaların çoğu gerçek dil kullanımını içermekte ve bugüne kadarki araştırmalar bu tür görevlerin üstesinden gelmede başarılı sonuçlar ortaya koymuşlardır (Abulaish vd., 2020). Ancak, söz sanatları için tam anlamıyla benzer bir başarı söz konusu değildir çünkü değişmeceli dili anlamak, insan zekâsının bile ayırt etmede zorlandığı bir durumdur (Harared ve Nurani, 2020; Kaya ve Alatas, 2022; Nagwanshi ve Madhavan, 2014; Savini ve Caragea, 2022). Bu sorunu en iyi temsil eden iletişimsel olgulardan biri ironi (Bettelli ve Panzeri, 2023; da Rocha Junqueira vd., 2023a; Frenda vd., 2023; Hamza vd., 2023; Maladry vd., 2023; Malik vd., 2023; OrtegaBueno vd., 2023; Saroj ve Pal, 2023; Tasnia vd., 2023; Tomás vd., 2023) ve iğneleme (Băroiu ve TrăuşanMatu, 2022; Goel vd., 2022; Govindan ve Balakrishnan, 2022; Misra, 2022; Pandey ve Singh, 2023; Qiao vd., 2023; Savini ve Caragea, 2022; D. K. Sharma vd., 2022; Vinoth ve Prabhavathy, 2022; Yue vd., 2023) tespitinde ortaya çıkmaktadır. İğneleme, bir metnin veya ifadenin gerçek anlamı ile kastedilen anlamı arasında bir tutarsızlık olduğunda ortaya çıkan bir ironi biçimidir. Bu tutarsızlık tipik olarak hoşlanmama, küçümseme veya aşağılama şeklinde kendini gösterir (Lad, 2023). İnsanların ürettiği ya da karşı karşıya kaldığı bu tür iğneleyici ifadelerin yer aldığı metinlerin verilerin önemli bir kısmı sosyal medya ortamlarında karşılarına çıkmaktadır (del Pilar Salas-Zárate vd., 2017; González-Ibáñez vd., 2011; Justo vd., 2014).

Çeşitli çalışmalar iğneleme tespiti için Twitter, Reddit ve Facebook'u kullanmıştır (Al-Ghadhban vd., 2018; da Rocha Junqueira vd., 2023b; Freitas vd., 2014; Ramteke vd., 2016). Facebook, Twitter, İnstagram, anketler vb. çevrimiçi sosyal ağlar alanında ironi ve iğneleme gibi davranışları tespit etmek, sosyal ve kişisel ilişkileri etkiledikleri için önemli bir sorumluluk haline gelmiştir (Ashwitha vd., 2020). Bu amaçla sosyal medya (Abulaish vd., 2020; Bouazizi ve Otsuki, 2016; Forslid ve Wikén, 2015), haberler (Nayak ve Bolla, 2022; Shrikhande vd., 2020; Zanchak vd., 2021), halka yönelik

konuşmalar (Zuhri ve Wahyudin Sagala, 2022) gibi farklı bağlamlarda ironi ve iğneleme gibi ifadeleri tespit etmek için kullanılabilecek etkili hesaplamalı yöntemler geliştirmek için önemli bir çaba sarf edilmiştir. Ancak siyasi içerik barındıran metinlerde ironi ve iğneleme gibi ifadelerin analizine yönelik araştırmaların oldukça az olduğu görülmüştür (Delmonte ve Stingo, 2016). Bu tür ifadelerin anlaşılması ve yorumlanması bağlamına bakılmaksızın iletişimin kalitesini etkilemesi bakımından önemlidir (Abulaish vd., 2020; AlGhadhban vd., 2018; Ashwitha vd., 2020; Bouazizi ve Otsuki, 2016; da Rocha Junqueira vd., 2023b; Forslid ve Wikén, 2015; Freitas vd., 2014). Bu doğrultuda sözlü ya da yazılı siyasi söylemlerde ironi ve iğneleme gibi hiciv içeren metinleri analiz eden yeni yaklaşımların ve araştırma çalışmalarının eksikliği hissedilmektedir (Mehndiratta ve Soni, 2019a).

Duygu analizi; karar verme, veri madenciliği, fikir madenciliği, bilgi yönetimi vb. alanlarda hayati öneme sahiptir ve telekom, tıbbi, çevresel, politik gibi birçok alanda ve farklı dillerde de yaygın olarak kullanılmaktadır (Abdar vd., 2020; Almuqren ve Cristea, 2021; Basiri vd., 2021; Kandasamy vd., 2020). Sosyal medya, bireylerin kişisel fikirleri diğer insanlara ifade etmek amacıyla çeşitli konulardaki fikirlerini gerçek zamanlı olarak paylaştıklarından, analiz edilmesi gereken büyük miktarda ham veri bulunmaktadır. Dolayısıyla bu verilerin manuel olarak sınıflandırılması imkânsız hale gelmektedir (Taşlıoğlu, 2014). Söylenen ile söylenmek istenen ifadelerin sosyal medyada yaygın olması ve makineler için tespit edilmesinin zor olması nedeniyle doğal dil işleme alanında büyük ilgi görmüştür (Yue vd., 2023). İroni ve iğneleme tespiti, DDİ'de önemli bir hesaplamalı işlem problemidir ve makineler ile insanlar arasındaki karşılıklı iletişim için bir arayüz olarak hizmet etmek üzere daha iyi anlaşılması için gereklidir (Ashwitha vd., 2020). İroninin tespiti, metin söz konusu olduğunda ilginç ancak zorlu bir görevdir çünkü eksik tonlama, ifade veya jestler nedeniyle gerçekte kastedilenle karşılaştırıldığında çelişkili bir unsura sahip olabilir. İroni barındıran metin komik veya kötü, eleştirel veya övgü içerebilir, ancak hepsi potansiyel olarak bir şekilde söylenenden farklı anlam iletir (Sekharan vd., 2018; Zhu ve Wang, 2020). İroni, örtük bir duygu durumudur ve metin düzeyinde daha iyi tespit için bağlam ve çok modluluk gibi ek bilgilere ihtiyaç duyar (Chauhan vd., 2022). Ayrıca literatürde yapılan çalışmaların genellikle İngilizce üzerinden yapılması, elde edilen sonuçların Türkçe gibi

farklı bir dil yapısına sahip (Özkan ve Kar, 2022; Taşlıoğlu, 2014) ve internet ortamındaki diğer dillerle kıyaslandığında kısıtlı kaynağı olan bir dil sınıflandırmasına (Baykara ve Güngör, 2023) girmektedir. Literatürde yer alan çalışmalar, araştırmaların veri kümelerine ve analiz yöntemlerine bağlı olarak (i) iğneleme, (ii) ironi, (iii) benzetme, (iv) metafor, (v) hiciv, (vi) mübalağa ve (vii) mizah olarak kategorize edilebilen kavramlara odaklanmışlardır (Abulaish vd., 2020).

İronik ve iğneleyici ve ifadelerin metinsel verilerdeki belirsizliği nedeniyle, tespit edilmesi zordur. İğneleme ve ironi tespitine yönelik araştırmalar on yılı aşkın bir süreyi kapsamakla birlikte, son zamanlarda ilerlemeler kaydedilmiştir. Yapılan çalışmalar genellikle İngilizce (Bolucu ve Can, 2020; Clavel Quintero ve Arco García, 2018; Dimovska vd., 2018; Gedela vd., 2023; Samonte vd., 2018; Tasnia vd., 2023; Yan vd., 2023), Arapça (Mohammed vd., 2020; Talafha vd., 2021), Çince (L. Liu vd., 2020; L. Zhang vd., 2022) ve diğer diller (Anan vd., 2023; Bharti vd., 2017; D. Jain vd., 2020; Lunando ve Purwarianti, 2013; Okimoto vd., 2021; Samonte vd., 2018; Y. Sharma ve Mandalam, 2021; Tasnia vd., 2023; Yunitasari vd., 2019) için yapılmış olup politik, ekonomik, Kültürel ve sosyal analizler sunan Türkçe köşe yazıları metinlerinde ironi ve iğneleme ifadelerine aynı anda odaklanan araştırmaya rastlanmamıştır. Yapılan çalışmalar veri kümesi açısından incelendiğinde Twitter, Reddit ve Amazon gibi internet siteleri üzerinden elde edilen kısa cümleler içeren veri kümelerinin kullanıldığı görülmektedir (Amir vd., 2016; Bamman ve Smith, 2021; Barbieri vd., 2014; Bharti vd., 2016; Joshi vd., 2016; Poria vd., 2016; Ptáček vd., 2014; Riloff vd., 2013; Schifanella vd., 2016). Bunun dışında yine internet platformları üzerinden elde edilen birkaç cümleden oluşan veri kümeleri de mevcuttur (Filatova, 2012; Son vd., 2019; Subramanian vd., 2019; Tay vd., 2018). İnternet platformları üzerinden elde edilen veri kümeleri genellikle etiket alanı kontrolü ile etiketlenmektedir. Bunun dışında elle etiketleme yöntemleri de mevcuttur. Yapılan çalışmalar metodoloji açısından incelendiğinde kural temelli yaklaşımlar (Bharti vd., 2015; Davidov vd., 2010; González-Ibáñez vd., 2011; Sundararajan ve Palanisamy, 2020), makine öğrenmesi yaklaşımları (Abercrombie ve Hovy, 2016; Bamman ve Smith, 2021; Jansi, 2018; Joshi vd., 2015; Mukherjee ve Bala, 2017; Parde ve Nielsen, 2018; Riloff vd., 2013; Thakur vd., 2020), derin öğrenme temelli yaklaşımlar (Barbieri vd., 2014; D. Ghosh vd., 2017;

A. Kumar vd., 2020b; Y. Liu vd., 2022; Mehndiratta ve Soni, 2019b; Poria vd., 2016; Salim vd., 2020), kolektif öğrenme yaklaşımları (Gupta vd., 2020; T. Jain vd., 2018; Potamias vd., 2019) ve dönüştürücü temelli yaklaşımlar (Gregory vd., 2020; Javdan ve MinaeiBidgoli, 2020; Kalaivani ve Thenmozhi, 2020; A. Kumar vd., 2020; Lan vd., 2020; Parameswaran vd., 2021; Srivastava vd., 2020; Zeng vd., 2019) kullanılmaktadır. Yapılan çalışmalar incelendiğinde veriler genellikle sosyal medyadan olduğu ve kısa cümleler içeren, etiket arama yöntemiyle elde edildiği anlaşılmaktadır. Kullanılan veri kümelerinin metodolojinin performansın iyi olarak görünmesine neden olacağı düşünülmektedir. Metodolojiler incelendiğinde dönüştürücü temelli yöntemlerin daha başarılı olduğu gözlemlenmektedir. Literatürde genellikle sınıflandırma üzerine çalışmalar yapılırken bağlamın hedef sınıf üzerindeki etkisi araştırılmamıştır.

Bharti ve arkadaşları (Bharti vd., 2016), Twitter sosyal medya ortamından elde edilen İngilizce tweetlerde iğneleyici duyguları tespit etmeye yönelik büyük veri tabanlı bir yaklaşım önermektedir. Çalışmada yaklaşık 1,45 milyon tweet içeren bir veri kümesi üzerinde Apache Hadoop çerçevesi kullanılarak iğneleme tespiti için önerilen yöntemler arasında Gizli Markov Modeli tabanlı sözcük etiketleme yöntemi, MapReduce tabanlı olan ayrıştırmaya dayalı sözlük oluşturma, ara sözcük başlangıcı ve zıt anlamlı kelime çifti ile olumlu duygu olmak üzere üç ana algoritma ve bunların birleşimi yer almaktadır. Performans değerlendirmesi için kesinlik, duyarlılık ve F1-skor metrikleri kullanılarak en iyi sonuç, bu üç yöntemin birleşimiyle elde edilmiştir. Birleşim yöntemi kesinlik, duyarlılık ve F1-skor metriklerinde sırasıyla %97, %98 ve %97 başarı değeri elde etmişlerdir.

Bouazizi ve Otsuki (Bouazizi ve Otsuki, 2016), Twitter üzerinden sarcasm etiketini kullanarak elde ettikleri 58 609 İngilizce tweet verisi ile iğneleme tespitine yönelik bir çalışma gerçekleştirmişlerdir. Çalışmada, verileri iğneleme ve değil olarak ikili sınıflandırmayı amaçlamaktadır. Çalışmada, duyguyla ilgili özellikler, noktalama işaretleriyle ilgili özellikler, söz dizimsel ve anlamsal özellikler ile desen tabanlı özellikler kullanılarak Rastgele Orman (RO), Destek Vektör Makineleri (SVM), k-En Yakın Komşu (kNN) ve Maksimum Entropi yöntemleri ile sınıflandırma işlemi

gerçekleştirilmiştir. En iyi performansı Rastgele Orman sınıflandırıcısı ile %83,1 doğruluk, %91,1 kesinlik, %73,4 duyarlılık ve %81,3 F1-skor değeri elde edilmiştir.

Ravi ve Ravi (Ravi ve Ravi, 2017), ironi ve iğneleme tespitine yönelik BURR (Burfoot ve Baldwin, 2009) veri kümesi ve kendileri web kazıma yöntemi ile elde ettikleri FAKE ve AMA isimli toplam 3 veri kümesi üzerinde gerçekleştirdikleri çalışmada metinlere ait özellikleri dil bilimsel, anlamsal, psikolojik ve unigram özellikleri çıkarmışlardır. Özellik seçimi için bilgi kazanımı, Chi-Kare ve korelasyon tabanlı yöntemler kullanılmıştır. Sınıflandırma için Destek Vektör Makineleri, Lojistik Model Ağaçları (LMT), Yapay Sinir Ağları, Lojistik Regresyon, Rastgele Orman, Sınıflandırma ve Regresyon Ağaçları, Karar Ağaçları, Tekrarlanan Aşamalı Budama ve Bayes Ağı algoritmaları uygulanmıştır. BURR veri kümesi en iyi performansı Lojistik Regresyon algoritması ile %85,86 F1-skoru, FAKE veri kümesinde Lojistik Model Ağaçları algoritması ile %92,50 F1-skoru, AMA veri kümesinde ise Rastgele Orman algoritması ile %95,20 F1-skoru elde edilmiştir.

Mukherjee ve Bala, (Mukherjee ve Bala, 2017), Twitter üzerinden #sarcasm ve #sarcastic etiketini kullanarak elde ettikleri 15 000 İngilizce tweet verisi ile iğneleme tespitine yönelik bir çalışma gerçekleştirmişlerdir. Çalışmada, çapraz doğrulama kullanarak verileri iğneleme ve değil olarak ikili sınıflandırmayı amaçlamaktadır. Çalışmada, içeriğe bağlı özellikler, işlevsel kelime özellikleri, sözcük tür etiketi özellikleri, ardışık dizi özelliklerini kullanılarak Naive Bayes ve Maksimum Entropi yöntemleri ile sınıflandırma işlemi gerçekleştirilmiştir. En iyi performansı Naive Bayes sınıflandırıcısı ile %73 doğruluk ve %76 F1-skor değeri elde edilmiştir.

Salas-Zárate ve arkadaşları (del Pilar Salas-Zárate vd., 2017), Twitter üzerinden web kazıma yöntemi ile elde ettikleri 10 000 Meksika ve İspanyol dillerine ait veri ile iğneleme tespitine yönelik bir çalışma gerçekleştirmişlerdir. Çalışmada, dil bilimsel araştırma ve kelime sayımı yöntemi üzerinden metinlerin özellikleri çıkarılıp, Destek Vektör Makineleri, Naive Bayes ve J48 yöntemleri ile sınıflandırma işlemi gerçekleştirilmiştir. En iyi performansı her iki veri kümesinde Destek Vektör Makineleri yöntemi Meksika dili için doğruluk, kesinlik, duyarlılık ve F1-skor

metrikleri için %85,50, İspanyolca için doğruluk, duyarlılık, F1-skor metrikleri için %84,00 ve kesinlik metriği için %84,60 değeri elde etmişlerdir.

Ghosh ve arkadaşları (D. Ghosh vd., 2018), Twitter, Reddit sosyal platformlarında etiketler vasıtasıyla topladıkları ve internet tartışma metinleri v2(IAC) (Walker vd., 2012) İngilizce veri kümeleriyle yaptıkları çalışmada iğneleme tespitine yönelik ikili sınıflandırma çalışması gerçekleştirmişlerdir. Özellik çıkarımı için konuşma birimi n-gramları, duygu sözlükleri ve bağlam tabanlı özellikler analiz edilmiştir. Sınıflandırma için Uzun-Kısa Süreli Bellek (LSTM) ağları, özellikle koşullu LSTM ve cümle düzeyinde dikkat (attention) mekanizması içeren modeller uygulanmıştır. En iyi performansı Reddit veri kümesinde cümle düzeyinde dikkat LSTM modeli ile %76,67 F1-skoru, Twitter veri kümesinde cümle düzeyinde dikkat LSTM modeli ile %76,36 F1-skoru, IAC veri kümesinde cümle düzeyinde dikkat LSTM modeli ile %73,70 F1-skoru ile elde edilmiştir.

Yunitasari ve arkadaşları (Yunitasari vd., 2019), Twitter sosyal platformundan etiketler vasıtasıyla toplanan Endonezce'ye ait 3 000 örnekten oluşan veri kümesiyle yaptıkları çalışmada iğneleme tespiti, duygu analizi ve iğneleme tespitinin duygu analizi üzerindeki etkisine yönelik çalışma gerçekleştirmişlerdir. İğneleme tespitinde özellik çıkarımı için duygu özellikleri, noktalama özellikleri, sözcüksel ve söz dizimsel özellikleri kullanırken, duygu analizinde Terim Sıklığı-Ters Belge Sıklığı (TF-IDF) özellik çıkarım yöntemi kullanılmıştır. Sınıflandırma yöntemi olarak iğneleme tespitinde Rastgele Orman, duygu analizinde Naive Bayes algoritmaları kullanılmıştır. İğneleme tespitinde %80,40 doğruluk, %83,20 kesinlik, %91,30 duyarlılık metrik değerleri elde edilirken, iğneleme tespiti kullanılarak duygu analizi doğruluğunda %5,49 oranında iyileşme elde edilmiştir.

Naz ve arkadaşları (Naz vd., 2019), Twitter, Kaggle ve thesarcasmdetector.com sosyal platformlarından elde ettikleri iğneleme içeren veri kümesi üzerinde Evrişimli Sinir Ağları yöntemi kullanarak iğneleme tespiti gerçekleştirmişlerdir. Özellik çıkarımında önceden eğitilmiş modeller yerine eğitim sırasında anlamsal açıdan benzer kelimelerin yakın olarak konumlandığı bir teknik kullanılmıştır. Yapılan çalışmada tüm verilerin

kullanıldığı durumda önerilen yöntem %97 kesinlik, %98 duyarlılık ve %98 F1-skoru değeri elde edilmiştir.

Son ve arkadaşları (Son vd., 2019) Twitter'dan SemEval 2015 Task 11 (A. Ghosh vd., 2015) veri kümesi ve yaklaşık 40 000 rastgele tweet içeren veri kümeleri üzerinde iğneleme tespiti gerçekleştirmişlerdir. Özellik çıkarımı için GloVe kullanılarak kelime gömmeleri oluşturulmuş, bu gömmeler çift yönlü LSTM (BiLSTM) ve yumuşak dikkat mekanizması ile anlamsal bağlam vektörlerine dönüştürülmüştür. Ayrıca, noktalama işaretleri gibi yardımcı özellikler evrişimli sinir ağı ile birleştirilmiştir. Önerilen sAttBiLSTM convNet isimli model ile %97,87 doğruluk değeri elde ederek, dikkat mekanizmalı BiLSTM model %89,03, dikkat mekanizması olmayan BiLSTM modelden %86,32 daha iyi sonuçlar elde etmiştir.

Bharti ve arkadaşları (Bharti vd., 2020), Hindistan'da en çok konuşulan ancak kısıtlı kaynaklara sahip Telugu diline ait televizyon şovlarından elde ettikleri soru-cevap şeklinde konuşma cümleleri içeren veri kümesi üzerinde ünlem, yoğunlaştırıcı, soru işareti ve ünlem sembolü gibi hiperbolik unsurların frekans ve dağılımları kullanarak iğneleme tespiti gerçekleştirmişlerdir. Çalışmada olumsuzluk kelimeleriyle başlayan cümleler için hazırlanmış olumsuz Telugu olumsuzluk kelimesi başlangıcı (TNWS), ünlem kelimeleriyle başlayan cümleler için Telugu ünlem kelimesi başlangıcı (TIWS) modeli ve cümlenin soru biçiminde olması durumuna göre çalışan soru biçimli yanıt (RiFoQ) modelleri önerilmektedir. Çalışmanın sonuçlarında TNWS yöntemi %91 doğruluk, %74,10 kesinlik, %69,69 duyarlılık ve %71,80 F1-skoru, TIWS yöntemi %88,50 doğruluk, %84 kesinlik, %85,10 duyarlılık ve %84,60 F1-skoru ve RiFoQ yöntemi %91,50 doğruluk, %65,30 kesinlik, %68 duyarlılık ve %66,60 F1-skoru değeri elde etmiştir.

Potamias ve arkadaşları (Potamias vd., 2020) iğneleme tespitine yönelik olarak dört farklı veri kümesi üzerinde kapsamlı bir çalışma gerçekleştirmiştir. Çalışmada SemEval2018 (Hee vd., 2018) ironik tweet veri kümesi, iğneleyici tweetlerin yer aldığı veri kümesi (Riloff vd., 2013), Reddit SARC2.0 politik yorum veri kümesi (Khodak vd., 2017) ve SemEval2015 Task 11 (A. Ghosh vd., 2015) veri kümeleri kullanılmıştır.

Önerilen yöntemde, önceden eğitilmiş güçlendirilmiş ve optimize edilmiş BERT yaklaşımı (RoBERTa) transformer ağı, BiLSTM katmanı ve tekrarlayan evrişimli sinir ağı (RCNN) ile birleştirerek kelimelerin bağlamsal temsillerini yakalayarak figüratif dili tespit etmeyi amaçlamaktadır. Önerilen RCNN-RoBERTa modeli; SemEval2018 veri kümesinde %82 doğruluk, %81 kesinlik, %80 duyarlılık ve %80 F1-skoru, ironik veri kümesinde %91 doğruluk, %90 kesinlik, %90 duyarlılık ve %90 F1-skoru, Reddit SARC2.0 veri kümesinde %79 doğruluk, %78 kesinlik, %78 duyarlılık ve %78 F1-skoru değerlerini elde etmiştir.

Sundararajan ve Palanisamy (Sundararajan ve Palanisamy, 2020) İngilizce twitter verileri üzerinde farklı iğneleme tiplerinin tespitini yapmışlardır. Çalışmada sosyal medyadan sarcasm, sarcastic etiketleri vasıtasıyla topladıkları 76 799 tweet içeren, nefret söylemi ve saldırgan dil etiketlerine sahip veri kümesi kullanılmıştır. Çalışmada, yoğunlaştırıcılar, duygu puanı, emoji duygu değeri ile özellik çıkarımı yapıldıktan sonra en uygun alt küme, önerilen çoklu kural tabanlı kolektif özellik seçimi yöntemiyle belirlenmiştir. Bulanık mantık ve kaba küme tabanlı çoklu kural yaklaşımı ile iğneleyici ifadeler kibar, kaba, öfkeli ve soğukkanlı olmak üzere dört türe ayrılmıştır. Önerilen model, alaycılık tespitinde %92,7 doğruluk; tür tespitinde ise kibar için %95,98, kaba için %96,20, öfkeli için %99,79 ve soğukkanlı için %86,61 doğruluk değeri elde etmiştir.

Meriem ve arkadaşları (Meriem vd., 2021) tarafından, sosyal ağlarda iğneleyici ifadelerin tespiti amacıyla metinsel içeriği ile yazarın geçmiş tweetleri, alıcı ile ilişki düzeyi ve cevaplarda yer alan ironi göstergeleri gibi sosyal bağlam özelliklerini kullanan Bulanık İğneleme Puanı Hesaplama (SSC) yaklaşımı önerilmiştir. Çalışmada metin, geçmiş, kullanıcı tanımlama ve cevap gibi dört ana özelliğin evrişimli sinir ağı (CNN) tabanlı skorlamayla ağırlıklandırılarak birleşmesini ve bulanık mantık ile iğneleyici, belirsiz, iğneleyici değil sınıflarına atanmasını içermektedir. Değerlendirmede SemEval 2014 Task 9 (Rosenthal vd., 2019) veri kümesi üzerinde yapılan çalışmada ikili sınıflandırmada en iyi sonuçlar iğneleme sınıfı için %88,44 duyarlılık, %84,1 kesinlik ve %86,21 F1-skoru, çok sınıflı bulanık sınıflandırmada ise

kesinlik değerleri her iki sınıf için yaklaşık %100'e ulaşmıştır. Bamman (Bamman ve Smith, 2021) veri kümesinde ise %82,50 doğruluk değeri elde edilmiştir.

Li ve arkadaşları (J. Li vd., 2021), Twitter ve Reddit sosyal platformlarından oluşturulan veri kümelerini kullanarak sağduyu bilgisi ile iğneleme tespiti gerçekleştirmişlerdir. Özellik çıkarımı için önceden eğitilmiş BERT modeli ile metin kodlaması yapılmış ve COMET modeli ile sağduyu bilgisi üretilmiştir ardından duygu kutupluluğu ve dikkat mekanizmasına dayalı seçim yöntemleriyle entegre edilmiştir. Önerilen model ile twitter1 (A. Ghosh ve Veale, 2016) veri kümesi üzerinde %85,52 kesinlik, %85,98 duyarlılık ve %85,73 F1-skoru, twitter2 (Ptáček vd., 2014) veri kümesi üzerinde %85,16 kesinlik, %86,05 duyarlılık ve %85,46 F1-skoru, Reddit veri kümesi (Khodak vd., 2017) üzerinde %76,10 kesinlik, %76,10 duyarlılık ve %76,10 F1-skoru elde edilmiştir. F1-skor metriğinde twitter1, twitter2 ve Reddit veri kümelerinde sırasıyla %3,56, %1,22 ve %3,18 iyileşme sağlamıştır.

Chia ve arkadaşları (Chia vd., 2021), sosyal platformundan etiketler vasıtasıyla elde edilen veri kümeleri üzerinde metin ön işleme ve TF-IDF özellik çıkarım yöntemlerini uyguladıktan sonra Naive Bayes, Destek Vektör Makineleri, hata azaltımı için tekrarlanan artırımlı budama (RIPPER) gibi makine öğrenmesi algoritmaları ile ironi ve iğneleme tespiti yapmışlardır. Çalışmada kullanılan SemEval2018 (Hee vd., 2018) veri kümesi üzerinde RIPPER algoritması ile veri ve etiket kullanımı ile %89,90, etiket kullanmadan %61,60 ve sadece etiket sınıflandırması ile %89,30 F1-skoru ile en iyi sonucu elde etmişlerdir.

Jamil ve arkadaşları (Jamil vd., 2021), Twitter ve haber başlıklarından oluşan veri kümesi üzerinde yaptıkları çalışmada, CNN ve LSTM yöntemlerini birleştirerek iğneleme tespiti gerçekleştirmişlerdir. Özellik çıkarımında CNN, model eğitimi ve test işlemlerinden LSTM kullanarak, makine öğrenmesi yöntemleri ve TF-IDF gibi kelime temsil yöntemleri ile karşılaştırmışlardır. Önerilen yöntem %92 model doğruluğu sağlayarak diğer yöntemlerden daha başarılı olmuştur.

Wen ve arkadaşları (Wen vd., 2022), Çince iğneleme tespiti yapmak amacıyla anlam birimi bilgisi ve yardımcı bağlam bilgisini birleştiren anlam birimi ve yardımcı geliştirilmiş dikkat sinir modeli (SAAG) isimli bir model önermektedirler. Çalışmada haber sitesinden toplanan ve 720 haber başlığı ile 4 972 yorumdan oluşan, GuanSarcasm (Gong vd., 2020) isimli Çince iğneleme veri kümesi kullanılmıştır. Yöntem olarak, kelime düzeyinde OpenHownet (Qi vd., 2019) modelinden elde edilen anlam birimi bilgisi Çift Yönlü Kapılı Tekrarlayan Birim (BiGRU) ve çok başlıklı dikkat mekanizmasıyla, cümle düzeyinde ise haber başlığı gibi yardımcı bilgiler yine kapılı tekrarlayan birim (GRU) ve öz-dikkat katmanıyla eğitim işlemi yapılmıştır. Ayrıca, dönüştürücü tabanlı çift yönlü kodlayıcı gösterimleri (BERT) tabanlı modellerde anlam birimi bilgisi ve yardımcı bilgilerinin birleştirilmesi ile oluşturulan $BERT^{SSAS}$ modeli ile çalışmalar gerçekleştirilmiştir. Deneysel sonuçlara göre SAAG modeli %72,07 doğruluk, %73,17 kesinlik, %73,05 duyarlılık ve %73,02 F1-skoru değeri, $BERT^{SSAS}$ modeli %76 doğruluk, %76,10 kesinlik, %75,91 duyarlılık ve %75,90 F1-skoru değeri elde etmiştir.

Jain ve arkadaşları (D. K. Jain vd., 2022), Hintçe tweetlerde iğneleme tespiti için LSTM ve SVM'yi birleştiren, kelime–emoji gömme temelli amalgam sinir mimarisi (TANA) modelini önermektedirler. Çalışmada veri kümesi olarak işaretli etiketlerden oluşturulmuş Hintçe 1 004 tweetten oluşan Sarc-H olarak isimlendirdikleri veri kümesi kullanılmıştır. Yöntem olarak fastText tabanlı Hintçe kelime temsilleri ile emoji2vec tabanlı emoji gömmeleri birleştirilerek vektör oluşturulmuş, bu vektör LSTM katmanında işlenmiş ve çıktı katmanında kare hinge kayıp fonksiyonu kullanan SVM ile sınıflandırma yapılmıştır. Çalışma sonucunda önerilen TANA modeli %97,01 doğruluk, %96,40 kesinlik, %97,10 duyarlılık ve %96,75 F1-skoru ile LSTM ve yapay sinir ağı tabanlı diğer yöntemlerden daha iyi sonuç elde etmiştir. Emojiler çıkarıldığında F1-skoru ve doğruluk metriklerinde düşüşler gözlemlenmiş olması emojilerin iğneleme tespitinde önemli katkısını göstermektedir.

Li ve arkadaşları (G. Li vd., 2022), İngilizce veri kümeleri üzerinde iğneleme tespitini amaçlayan bir çalışma gerçekleştirmişlerdir. Çalışmada, bağlamsal bilgiyi BiLSTM kodlayıcı ile elde ettikten sonra, bağımlılık ilişkilerini modelleyen İlişkisel Grafik

Dikkat Ağı (RGAT) ve duygusal tutarsızlıkları yakalamak için Grafik Evrişimsel Ağ (GCN) birlikte kullanılmasıyla ortaya çıkan ARGAT isimli model ile iğneleme tespiti amaçlanmaktadır. Çalışmada veri kümesi olarak IACV12 (Lukin ve Walker, 2017), Twitter1 (Riloff vd., 2013), Twitter2 (Ptáček vd., 2014) ve Reddit kaynaklı filmler ve teknoloji (Khodak vd., 2017) veri kümeleri kullanılmıştır. Çalışma sonucunda önerilen ARGAT modeli IACV1 veri kümesinde %72,25 doğruluk, %72,26 kesinlik, %72,26 duyarlılık ve %72,26 F1-skoru, IACV2 veri kümesinde %78,22 doğruluk, %78,41 kesinlik, %78,19 duyarlılık ve %78,21 F1-skoru, Twitter1 veri kümesinde %85,81 doğruluk, %83,19 kesinlik, %76,24 duyarlılık ve %79,78 F1-skoru, Twitter2 veri kümesinde %84,28 doğruluk, %84,28 kesinlik, %84,28 duyarlılık ve %84,28 F1-skoru, filmler veri kümesinde %75,82 doğruluk, %75,82 kesinlik, %75,82 duyarlılık ve %75,82 F1-skoru, teknoloji veri kümesinde %76,13 doğruluk, %76,13 kesinlik, %76,13 duyarlılık ve %76,13 F1-skoru değerleri elde etmiştir.

Chauhan ve arkadaşları (Chauhan vd., 2022), iğneleme tespiti ve duygu analizi geliştirmek amacıyla MUStARD (Castro vd., 2019) veri kümesi üzerinde SEEmoji isimli yöntem önermiştir. Veri kümesinde her bir konuşma ifadesi için ilgili emoji, emojinin duygu ve duygu durumu bulunmaktadır. Çalışmada metin, görsel, akustik gibi farklı girdiler kullanılarak, iğneleme tespiti birincil görev; duygu ve duygu durumu tespiti ise ikincil görevler olarak ele alınan çok görevli derin öğrenme çerçevesi önerilmiştir. Yöntem olarak emojileri algılayan çok modlu dikkat sistemi (GEMA) mekanizması geliştirilmiş; metinler için fastText ve BiGRU, görseller için ortalama çerçeve özellikleri, akustik için ortalama özellikler, emojiler için emoji2vec gömme yöntemleri kullanılmıştır. Çalışma sonucunda, önerilen model %77,9 kesinlik, %76,9 duyarlılık ve %76,7 F1-skoru değerleri ile mevcut yöntemlerden daha yüksek performans göstermiştir.

Savini ve Caragea (Savini ve Caragea, 2022), İngilizce metinlerde iğneleme tespitinde BERT tabanlı dil modelleri ile ara görev aktarım öğrenmesini kullanarak. Çalışmada hedef görev olarak iğneleme tespiti belirlenmiş, ara görevler olarak ise EmoNet veri kümesinden elde edilen ince taneli duygu tespiti, EmoNetSent veri kümesinden kutuplaşmış duygu tespiti ve IMDB film yorumları veri kümesi duygu sınıflandırması

için kullanılmıştır. Hedef görev için üç farklı veri kümesi kullanılarak ince ayar işlemi gerçekleştirildikten sonra iğneleme tespiti için kullanılmıştır. Çalışmada kullanılan SARC (Khodak vd., 2017) veri kümesinde BERT + TransferEmoNetSent modeli %77,53 F1-skoru, Sarcasm V2 (Oraby vd., 2017) veri kümesinde BERT + TransferIMDB modeli %80,55 F1-skoru, SARCTwitter (Mishra vd., 2016) veri kümesinde BERT + TransferEmoNetSent modeli %97,43 F1-skoru elde etmiştir. Çalışma çıktıları, küçük veri kümelerinde ara görev aktarımının, özellikle duygu bilgisinin, iğneleme tespitinde anlamlı iyileşme sağladığını göstermektedir.

Misra ve Arora (Misra ve Arora, 2023), haber başlıklarından oluşan ve kendilerinin hazırladığı veri kümesi kullanarak haber başlıklarında iğneleme tespiti gerçekleştirmişlerdir. Veri kümesine ait özellik çıkarımından word2vec kelime gömme yöntemi kullanarak metinlerin temsillerini oluşturduktan sonra CNN, LSTM ve dikkat mekanizmasını kullanarak oluşturdukları hibrit model ile sınıflandırma işlemi gerçekleştirilmiştir. Önerilen yöntem ile test veri kümesi olarak ayırdıkları %10 örnek ile %89,70 model doğruluğu elde edilmiştir.

Yao ve arkadaşları (Yao vd., 2023), Twitterdan elde edilen ve metin, görsel, görsel içindeki metin ve resim başlığı bilgilerinden oluşan İngilizce veri kümesini (Cai vd., 2019) kullandıkları çalışmalarında, çok modlu, çok etkileşimli ve çok hiyerarşik bir sinir ağı (M3N2) yöntemi ile iğneleme tespiti gerçekleştirmişlerdir. Önerilen yöntemin özellik çıkarımında tekli, ikili, üçlü, dörtlü mod etkileşimlerini, kapı mekanizması, yönlendirilmiş dikkat, çok adımlı işlem, öz-dikkat ve dikkat havuzlama teknikleri kullanılmıştır. Çalışma sonucunda %83,23 doğruluk, %77,02 kesinlik, %82,48 duyarlılık ve %79,66 F1-skor metrik sonuçlarını elde ederek, doğruluk, kesinlik ve F1-skor metriklerinde aynı veri kümesi üzerinde yapılan çalışmaları geride bırakmıştır.

Tasnia ve arkadaşları (Tasnia vd., 2023), İngilizce ve İspanyolca dillerinde sosyal medya gönderilerinde mizah ve ironi tespiti amacıyla önerdikleri yöntemde GloVe, ELMo, BERT ve Flair gibi önceden eğitilmiş kelime gömme modellerini yığmalı olarak birleştirilmiş, ardından LSTM tabanlı belge vektörü oluşturulmuş ve çok katmanlı doğrusal sınıflandırma mimarisi hedef sınıf tahmini gerçekleştirilmiştir. Çalışmada,

SemEval2021 (Meaney vd., 2021) , SemEval2018 (Hee vd., 2018) ve IberLEF (Chiruzzo vd., 2019) ve IberEval (Castro vd., 2018) veri kümesi üzerinde gerçekleştirilmiştir. Yöntem, SemEval2021 veri kümesinde %92,20 doğruluk, %93,50 kesinlik, %93,80 duyarlılık ve %93,60 F1-skor, SemEval2018 veri kümesinde %71,40 doğruluk, %59,30 kesinlik, %89,40 duyarlılık ve %71,20 F1-skor, IberLEF veri kümesinde %84,90 doğruluk, %80,70 kesinlik, %80,80 duyarlılık ve %80,70 F1-skor ve IberEval veri kümesinde %86,60 doğruluk, %82,70 kesinlik, %81,40 duyarlılık ve %82 F1-skor metrik değerleri elde etmiştir.

Pandey ve Singh (Pandey ve Singh, 2023), Hintçe ve İngilizce metinleri kullanarak sosyal medya paylaşımlarında iğneleme tespitini amaçlayan bir çalışmada 5 250 örnek bulunan veri kümesi (Swami vd., 2018) kullanılarak gerçekleştirmiştir. Önerilen yöntemde, çok dilli önceden eğitilmiş BERT modeli ile elde edilen gömme vektörleri ile LSTM ağının kullanılması sonucu ikili sınıflandırma işlemi gerçekleştirilmiştir. Önerilen modelin veri kümesi üzerinde elde ettiği performans metrikleri, Naive Bayes, lojistik regresyon, SVM, karar ağacı, rastgele orman, en yakın komşu (KNN), gradyan artırma gibi makine öğrenmesi yöntemleri ve derin sinir ağları (DNN), CNN, LSTM ve BERT yöntemleri ile karşılaştırılmıştır. Çalışma sonuçları, BERTLSTM modelinin iğneleme sınıfı için %86 kesinlik, %99 duyarlılık, %92 F1-skoru; normal sınıfı için %100 kesinlik, %98 duyarlılık, %99 F1-skoru ile karşılaştırılan diğer modellere göre en yüksek performansı gösterdiğini ortaya koymaktadır.

Yue ve arkadaşları (Yue vd., 2023), İngilizce çok modlu iğneleme tespitini amaçlayan bir çalışmada Cai ve arkadaşları (Cai vd., 2019) tarafından oluşturulan ve İngilizce tweetlerden oluşan, metin, görsel ve görsel öznitelik bilgilerinin yer aldığı veri kümesi kullanılmıştır. Önerilen yöntem olan KnowleNet modeli, BERT ile metin ve görsel açıklama, ResNet152 ile görsel özellik çıkarımı yaparak; kelime düzeyinde ve örnek düzeyinde anlamsal benzerlik tespitinin ardından triplet kayıp fonksiyonu içeren karşıt öğrenme yöntemiyle iğneleyici ve iğneleyici olmayan örneklerin tespit edilmesi hedeflenmiştir. Model, %88,87 doğruluk, %88,59 kesinlik, %84,18 duyarlılık ve %86,33 F1 puanı ile doğruluk, kesinlik ve F1-skoru metriklerinde en iyi sonucu elde etmiştir.

Grover ve Banati (Grover ve Banati, 2024), emojilerin çevrim içi metinlerde iğneleme tespitindeki rolünü inceledikleri çalışmalarında 29 377 metinden oluşan ve tamamı emoji içeren SarcOji (Grover ve Banati, 2022) veri kümesi ile birini twitter üzerinden oluşturdukları SarcOjiTest2 ve SarcOjiTest1 (Ptáček vd., 2014; Riloff vd., 2013; Shmueli vd., 2020) veri kümesi kullanılmıştır. Kelime gömme vektörü olarak GloVe, modellere girdi olarak sadece metinlerin yer aldığı PlainText, hem metinlerin hem emojilerin kullanıldığı TextWithEmojis son olarak en sık kullanılan emojiler ve metinlerin yer aldığı yöntemler, mimari olarak LSTM, BiLSTM, Bahdanau (Bahdanau vd., 2014) eklemeli dikkat ve Luong (Luong vd., 2015) çarpımsal dikkat gibi derin öğrenme yöntemleri ile eğitilmiştir. Çalışma sonucunda, SarcOjiTest1 veri kümesinde TextMaxEmoji girdi yönteminin çarpımsal dikkat mimarisi ile %52,32 F1-skoru, SarcOjiTest2 veri kümesinde eklemeli dikkat mimarisi ile %54,95 F1-skoru elde ederek emojilerin modele eklenmesinin iğneleme tespitini belirgin şekilde iyileştirdiği belirtilmektedir.

Helal ve arkadaşları (Helal vd., 2024), bağlam bilgilerini kullanarak iğneleme tespitini geliştirmek için RoBERTa ve damıtılmış BERT (DistilBERT) modelleri, MUStARD (Castro vd., 2019) ve haber başlıkları (Misra ve Arora, 2023) olmak üzere iki farklı veri kümesi üzerinde bağlam verileri eklenerek ince ayar ile eğitilmiştir. MUStARD veri kümesinde sahne diyalogları, konuşmacı bilgileri ve örtük duygular eklenerek doğruluk %89, F1 skoru %90'a ulaşmıştır. Haber başlıkları veri kümesinde açıklama, yazar adı ve haber bölümü gibi meta veriler eklenerek doğruluk %99 ve F1 skoru %99,7'ye ulaşılmıştır. Önerilen model, Reddit (SARC) veri kümesi üzerinde bağlam verisi olmadan %49,5 doğruluk ve %55 F1-skoru elde ederken bağlam eklenmesiyle doğruluk %60,8'e, F1-skoru ise %75'e yükselmiştir. Sonuçlar, bağlam bilgisinin iğneleme tespitinde belirgin performans artışı sağladığını göstermektedir. Ayrıca bağlamın daha kısa ve öz olarak özetlenmesinin eğitim süresinde %35,5 oranında azalmaya neden olduğu görülmektedir.

Liu ve arkadaşları (H. Liu vd., 2024), İngilizce dilinde çok modlu iğneleme tespitini amaçlayan çalışmalarında çok modlu duygu algılama özellikli hiyerarşik füzyon ağı (SAHFN) adlı bir model önermektedir. Çalışmada, metin, görüntü ve görüntüden elde

edilen öznitelik modaliteleri birlikte kullanılarak, farklı türde verilerden BERT, RoBERTa ve DenseNet121 ile özellik çıkarımı yapıldıktan sonra SenticNet7 (Cambria vd., 2022) modeliyle elde edilen kelime düzeyinde duygu puanlarıyla duyguya duyarlı dikkat mekanizması entegre edilmiştir. Farklı türde veriler arasında ilişkiler çapraz dönüştürücü katmanlarıyla modellenmiş ve duyguya duyarlı karşıt kayıp fonksiyonu kullanılmıştır. Çalışmada, Twitter çok çeşitli iğneleme veri kümesi (Cai vd., 2019) ve Multibully (Maity vd., 2022) veri kümesi kullanılmış olup, Twitter veri kümesinde RoBERTa ile %96,17 doğruluk, %95,74 kesinlik, %95,84 duyarlılık ve %95,79 F1-skoru, BERT ile %87,22 doğruluk, %82,71 kesinlik, %87,33 duyarlılık ve %84,95 F1-skoru; Multibully veri kümesinde ise RoBERTa ile %61,59 doğruluk ve %67,73 F1-skoru, BERT ile %59,87 doğruluk ve %63,30 F1-skoru elde edilerek mevcut yöntemleri geride bırakmıştır.

Zhong ve arkadaşları (Zhong vd., 2024), çok ölçekli ve çok aralıklı token düzeyinde metinsel ve görsel bilgileri modelleyerek görüntü-metin uyumunu ele almak için bir anlamsal geliştirme çerçevesi (SEF) önerilmektedir. Çalışmada, metinler önceden eğitilmiş BERT modeliyle, görseller ise görüntü dönüştürücüsü (Vision Transformer) ile kodlanarak çoklu ölçek ve çoklu kapsam düzeyinde metin–görüntü uyumu modellenmiştir. Veri türlerinin birbirleri ile etkileşimine küçük parça düzeyinde, büyük bütünlere ve genel düzeyde uyuma bakılarak karşıt öğrenme yöntemleri kullanılmıştır. Bunun yanında, Intra-modal çizge sinir ağları (GNN) ile cümle içindeki anlam ilişkileri belirlenmiş ve denetimli karşıt öğrenme sayesinde metinlerin anlamı daha güçlü hale getirilmiştir. Önerilen SEF yöntemi Twitter tabanlı çok modlu iğneleme veri kümesi (Cai vd., 2019) üzerinde test edilmiş ve %88,45 doğruluk, %85,35 kesinlik, %86,58 duyarlılık ve %85,96 F1-skor değeri ile en yakın yöntem olan HKE (H. Liu vd., 2022) modeline kıyasla doğrulukta %1, F1-skorunda %1,87 artış sağlamıştır.

Chakravarthi (Chakravarthi, 2025), Tamilce ve Malayalamca olmak üzere iki dravid dili için YouTube yorumlarından oluşan, elle etiketlenmiş yeni bir iğneleme veri kümesi geliştirmiştir. Yöntem olarak, makine öğrenmesi tabanlı modellerde TF-IDF özellik çıkarımı ile Naive Bayes, SVM, Lojistik Regresyon, Karar Ağaçları ve Rastgele Orman; ayrıca önceden eğitilmiş dönüştürücü tabanlı modeller olarak çoklu dil destekli BERT

(mBERT), hindi dilleri için çoklu dil destekli BERT (MuRIL), TamBERT ve MalBERT kullanılmıştır. Deneylerde veriler eğitim, geliştirme ve test kümelerine ayrılmış; hem tek dilli hem de kod karışımlı (codemixed) içerikler değerlendirilmiştir. Performans metriklerine göre, Tamil veri kümesinde en iyi sonuç TamBERT modeliyle %79,80 doğruluk, %74,30 kesinlik, %73,40 duyarlılık ve %73,80 F1-skor, Malayalam veri kümesinde ise en iyi sonuç MuRIL modeliyle %85,20 doğruluk ve %73,30 F1-skor olarak elde edilmiştir. Çalışma, Tamilce ve Malayalamca iğneleme tespitinde dillere özel ince ayar yapılmış dönüştürücü modellerin, çok dilli modellere kıyasla daha iyi performans gösterebileceğini ortaya koymaktadır.

Li ve arkadaşları (Q. Li vd., 2025), İngilizce dili için çoklu ortam verilerinde iğneleme tespiti yapmak amacıyla Vicuna7B tabanlı büyük dil modeli üzerine inşa edilmiş ve dört ana bileşeni olan görev odaklı girdi şablonu üretimi, bilgi getirme destekli üretim ile bağlamsal bilgi ekleme, zincirleme düşünme ile adım adım akıl yürütme ve metin-görüntü bilgilerini bütünleştiren bağlam füzyonu olan SarcasmGPT isimli model önermişlerdir. Çalışmada yaygın olarak kullanılan MUStARD (Castro vd., 2019) ve MMSD2.0 (Qin vd., 2023) iğneleme veri kümleri kullanılmıştır. SarcasmGPT, MUStARD veri kümesinde %82,31 doğruluk, %80,13 kesinlik, %84,37 duyarlılık ve %82,20 F1-skor, MMSD2.0 veri kümesinde ise %86,78 doğruluk ve %86,47 F1-skor elde ederek önceki derin öğrenme ve LLM tabanlı yaklaşımları geride bırakmıştır.

Sukhavasi ve Dondeti (Sukhavasi ve Dondeti, 2025), metin, görüntü ve ses verilerde iğneleme tespiti amacıyla hibrit EnsembleNet yaklaşımı önermişlerdir. Çalışmada görüntü özellikleri ResNet152 ile, metin özellikleri terim sıklığı ters ortalama belge Sıklığı (TFIADF) yöntemiyle çıkarılıp BiLSTM ağına verilerek, ses özellikleri ise mel frekanslı kepstral katsayıları (MFCC), tepe eğimi ve perde takibi gibi akustik öznitelikler kullanılarak VGG16 modeline aktarılmıştır. Üç farklı türden elde edilen özellikler, ağırlıklı birleştirme modu (WFM) ile birleştirilmiş ve sınıflandırma softmax katmanı ile yapılmıştır. Modelin hiper parametreleri, karmaşıklığı azaltmak ve performansı artırmak amacıyla tent kaotik snack optimizasyonu (TCSO) ile ayarlanmıştır. Çalışmada Memotion 7k (C. Sharma vd., 2020) ve MUStARD (Castro vd., 2019) veri kümeleri kullanılmış; Memotion 7k'da %98,64 doğruluk, %99,56

kesinlik, %99,45 duyarlılık ve %99,57 F1; MUStARD'da ise %98,99 doğruluk, %98,97 kesinlik, %98,78 duyarlılık ve %98,89 F1 skorları elde edilmiştir. Çalışma İngilizce dilindeki veri kümeleri üzerinde gerçekleştirilmiştir.

Zhang ve arkadaşları (Y. Zhang vd., 2025), metin ve görüntü içeren verilerde iğneleme tespitinde görüntü ve metin arasındaki anlamsal farkı azaltmak amacıyla aşamalı etkileşim modeli (PIMMS) önermektedir. Çalışmada önerilen yöntemde, metin verilerine ait özelliklerin RoBERTa ve görüntülere ait özelliklerin görüntü dönüştürücü model ile özelliklerinin çıkarılmasından sonra köprü katmanı kullanılarak metin ve görsel özellikler arasında ön etkileşim sağlanmış, ardından karşıt öğrenme tabanlı anlamsal hizalama ve çift akışlı dönüştürücü mimarisi ile çok modlu etkileşim gerçekleştirilmiştir. Son olarak, iğnelemeye ait bulguları birleştiren füzyon katmanı ile nihai sınıflandırma yapılmıştır. Önerilen model, Twitter'dan elde edilen İngilizce diline ait çok modlu iğneleme tespiti (Cai vd., 2019) veri kümesi üzerinde test edilmiştir. Önerilen PIMMS modeli, metin ve görüntülerden oluşan veri kümesinde %90,34 doğruluk, %89,63 kesinlik, %90,35 duyarlılık ve %89,91 F1-skor değeri ile aynı veri kümesi üzerinde test edilmiş modellerden daha iyi sonuç vermiştir. Önerilen model ayrıca Multibully (Maity vd., 2022) veri kümesi üzerinden test edilmiş ve %64,96 doğruluk, %64,74 kesinlik, %63,69 duyarlılık ve %64,21 F1-skor değeri elde etmiştir.

Türkçe için yapılan çalışmalar incelendiğinde, ironi ve hiciv üzerine olduğu ve ikili sınıflandırma kullanıldığı görülmektedir. Bölücü ve Can (Bölücü ve Can., 2025), sosyal medya kullanıcıların duygu ve düşüncelerinde ironi içeren ifadelerin tespiti için yaptıkları çalışmada, evrensel kavramsal bilişsel açıklama ve Çizge Sinir Ağları yöntemlerini birlikte kullanarak Türkçe ironi veri kümesi üzerinde %94,85 F1-skoru elde ederek en iyi modele göre %7,3 artış elde etmişlerdir. Aynı yöntem kullanılarak IronyTR veri kümesi üzerinde %72,83 doğruluk değeri elde ederek, en iyi modele göre %4,39 artış elde etmişlerdir. Öztürk ve arkadaşları (Öztürk vd., 2021), geleneksel makine öğrenmesi, derin öğrenme ve dönüştürücü modeller kullanarak yaptıkları ironi sınıflandırma çalışması gerçekleştirmişlerdir. Twitter üzerinden toplanan 600 örnekli veri kümesi üzerinde BERT yöntemini kullanarak %69,00 doğruluk ve %68,43 F1-skor değeri elde etmişlerdir. Karabaş ve Diri (Karabaş ve Diri, 2020), Twitter üzerinden elde

edilen ironi veri kümesi üzerinde sözcük torbası modeli ve önceden eğitilmiş Türkçe kelime gömmeleri, makine öğrenmesi yöntemleri ve derin öğrenme yöntemleri kullanarak ironi içeren ifadeleri sınıflandırma çalışmasında dikkat mekanizmasına sahip BiLSTM yöntemi kullanarak %84,30 doğruluk ve %82,80 F1-skoru elde etmişlerdir. Onan ve Toçoğlu (Onan ve Toçoğlu, 2020), Türkçe haber makalelerinde hiciv haberlerini tespit etmek için dilbilimsel ve psikolojik özellik kümeleri kullanarak, topluluk, derin öğrenme ve makine öğrenmesi yöntemleri ile sınıflandırma yapmışlardır. Elde edilen sonuçlara göre, rastgele orman algoritmasının Türkçede hiciv tespiti için %96,92 sınıflandırma doğruluğu ile en yüksek performansı sağladığı görülmektedir. Derin öğrenme tabanlı mimariler için, dikkat mekanizmasına sahip tekrarlayan sinir ağı mimarisi ile %97,72 sınıflandırma doğruluğu elde edilmiştir. Kara ve Karagöz (Kara ve Karagöz, 2025), ironi tespiti için yaptıkları çalışmada veri artırımın etkisi incelenmiş ve performans artırmak için kolektif öğrenme yöntemi kullanışlardır. Klasik makine öğrenmesi yöntemlerinden SVM ile %61 doğruluk, dönüştürücü modellerle %76,70 doğruluk ve kolektif öğrenme yöntemi ile %71 doğruluk değeri elde etmişlerdir. Öztürk (Öztürk,2024), ironi (IronyTR) ve hiciv (Kaggle+Zaytung) veri kümeleri üzerinde ayrı ayrı yaptığı çalışmada klasik makine öğrenmesi yöntemleri ve dönüştürücü yöntemlerden BERT'i kullanarak ikili sınıflandırma yapmıştır. İroni veri kümesinde makine öğrenmesi yöntemlerinden SVM ile %63,33 doğruluk ve %61,03 F1-skor en yüksek değeri elde ederken, BERT yöntemi ile %69,00 doğruluk ve %68,43 F1-skor değeri elde etmişlerdir. Hiciv veri kümesinde makine öğrenmesi yöntemlerinden Karar Ağaçları ile %91,16 doğruluk ve %91,04 F1-skor en yüksek değeri elde etmiştir. Cekinel ve Karagöz (Cekinel ve Karagöz, 2022), ironi ve iğnelemenin tespitinde istatiksel bağlantılar yerine nedensel çıkarım yöntemlerinin sonuçlarını analiz etmişlerdir. Çalışmada yanlılığın azaltılması amacıyla dilsel özelliklerin sonuçlara etki etmesinde neden sonuç ilişkisine dayanması incelenmiştir. IronyTR veri seti üzerinde ünlem noktalama işaretlerinin ironi ifadelerine neden olduğu belirtilmiştir. Reddit veri seti üzerinde BERT kullanan vekil sınıflandırıcı ile yapılan analizlerde %78,60 doğruluk ve %80,60 F1-skor değeri elde etmişlerdir. Büyükbaş ve arkadaşlarının (Büyükbaş vd., 2021) yaptıkları çalışmada ironik cümlelerin açıklanabilirliği için Shapley Katkı Açıklamaları (SHAP) ve Yerel Yorumlanabilir Model-bağımsız Açıklamalar (LIME) yöntemlerini kullanmışlardır. Çalışma sonucunda LSTM,

BiLSTM, BERT ve T5 yöntemleri kullanılarak elde edilen sonuçlarda BERT yöntemi %69,00 doğruluk ve %68,43 F1-skor elde ederek en iyi sonucu elde etmiştir. Taşlıoğlu ve Karagöz (Taşlıoğlu ve Karagöz, 2017), az sayıda özellik kullanarak ironi içeren ifadelerin tespitini denetimli öğrenme algoritmaları kullanarak gerçekleştirmiştir. Çalışmada özellik çıkarımı olarak köşeli parantez soru işareti, ünlem işareti, soru işareti, tırnak işareti, nokta, gülme emojisi, kızma emojisi, duygu puanları arasındaki fark, küçültme formları, nitelemeler, büyük harf kullanımına göre yapılmıştır. Özellik çıkarımından sonra bilgi kazancı yöntemi ile en iyi 5 özellik seçilerek sınıflandırma işlemi yapılmıştır. Rastgele Orman algoritması, %88,60 F1-skor değeri ile en iyi sonucu elde etmiştir. Cemek ve arkdaşlarının (Cemek vd., 2020), ironi içeren cümlelerin tespitini resmi olmayan metinler üzerinden sinirsel ağ tabanlı modeller ile gerçekleştirmişlerdir. Yöntem olarak LSTM, BiLSTM, BERT, SVM, Naive Bayes, veri kümesi olarak Twitter gönderileri kullanılmış ve en iyi sonucu, %87,50 doğruluk ve %87,23 F1-skor değerleri ile BERT elde etmiştir. Toçoğlu ve Onan (Toçoğlu ve Onan, 2019), hiciv içeren ifadeleri tespit etmek amacıyla haber sitelerinden hazırladıkları veri kümesi üzerinde n-gram ve TF-IDF yöntemleri elde ettikleri özellik vektörlerini SVM, Lojistik Regresyon, C4.5 ve Naive Bayes yöntemleri ile sınıflandırmışlardır. En iyi sonucu %89,70 doğruluk ile unigram ve TF özellik vektörü ile SVM yöntemleri elde etmiştir. Dülger (Dülger, 2018), Türkçe metinlerde ironi tespiti yaptığı çalışmasında desen tabanlı özellikler ve polarite tabanlı özellikler kullanarak özellik vektörü oluşturarak denetimli makine öğrenmesi algoritmaları ile sınıflandırmıştır. Çok katmanlı algılayıcı yöntemi %88,19 doğruluk ve %88,20 F1-skor elde ederek en iyi sonucu vermiştir.

2015-2025 yılları arasında makine öğrenmesi algoritmaları, derin öğrenme algoritmaları, ön eğitimli dil modelleri kullanılarak ironi veya iğneleme tespiti amacıyla gerçekleştirilen çalışmalar Çizelge 2.1'de sunulmuştur.

Çizelge 2.1. İğneleme ve ironi tespiti için gerçekleştirilen çalışmalarda kullanılan yöntemler, sonuçlar ve kullanılan veri kümeleri

Kaynak	Dil	Yöntem	Metrik (%)	Veri Kümesi
Bharti vd., 2016	İngilizce	Hibrit	Doğruluk: 97,00	Kendi Veri Kümesi
Bouazizi ve Otsuki, 2016	İngilizce	RF	Doğruluk: 83,1	Kendi Veri Kümesi
Ravi ve Ravi, 2017	İngilizce	LR	F1-skor: 85,86	BURR
Ravi ve Ravi, 2017	İngilizce	RF	F1-skor: 95,20	Kendi Veri Kümesi
Ravi ve Ravi, 2017	İngilizce	LMT	F1-skor: 92,50	Kendi Veri Kümesi
Mukherjee ve Bala, 2017	İngilizce	Naive Bayes	Doğruluk: 73,00	Kendi Veri Kümesi
del Pilar Salas-Zárate vd., 2017	İspanyolca	SVM	Doğruluk: 85,5	Kendi Veri Kümesi
del Pilar Salas-Zárate vd., 2017	Meksika	SVM	Doğruluk: 84,00	Kendi Veri Kümesi
D. Ghosh vd., 2018	İngilizce	LSTM	F1-skor: 76,67	Kendi Veri Kümesi
D. Ghosh vd., 2018	İngilizce	LSTM	F1-skor: 76,36	Kendi Veri Kümesi
D. Ghosh vd., 2018	İngilizce	LSTM	F1-skor: 73,70	IAC
Yunitasari vd., 2019	Endonezce	RF	Doğruluk: 80,40	Kendi Veri Kümesi
Naz vd., 2019	İngilizce	CNN	F1-skor: 98,00	Kendi Veri Kümesi
Son vd., 2019	İngilizce	sAtt-BiLSTM+ CNN	Doğruluk: 97,87	SemEval 2015 Task 11
Bharti vd., 2020	Telugu	TNWS	Doğruluk: 91,00	Kendi Veri Kümesi
Bharti vd., 2020	Telugu	TIWS	Doğruluk: 88,5	Kendi Veri Kümesi
Bharti vd., 2020	Telugu	RiFoQ	Doğruluk: 91,50	Kendi Veri Kümesi
Potamias vd., 2020	İngilizce	RCNN-RoBERTa	Doğruluk: 82,00	SemEval-2018
Sundararajan ve Palanisamy, 2020	İngilizce	Bulanık mantık	Doğruluk: 92,7	Kendi Veri Kümesi
Meriem vd., 2021	İngilizce	SSC	F1-skor: 86,21	SemEval-2014
J. Li vd., 2021	İngilizce	COMET	F1-skor: 85,73	Hazır Veri Kümesi
Chia vd., 2021	İngilizce	RIPPER	F1-skor: 89,90	SemEval-2018
Jamil vd., 2021	İngilizce	CNN + LSTM	Doğruluk: 92,00	Kendi Veri Kümesi
Wen vd., 2022	Çince	SAAG	Doğruluk: 72,07	GuanSarcasm
Wen vd., 2022	Çince	BERT-SSAS	Doğruluk: 76,00	GuanSarcasm
D. K. Jain vd., 2022	Hintçe	TANA	Doğruluk: 97,01	SarcH

Çizelge 2.1. İğneleme ve ironi tespiti için gerçekleştirilen çalışmalarda kullanılan yöntemler, sonuçlar ve kullanılan veri kümeleri (devam)

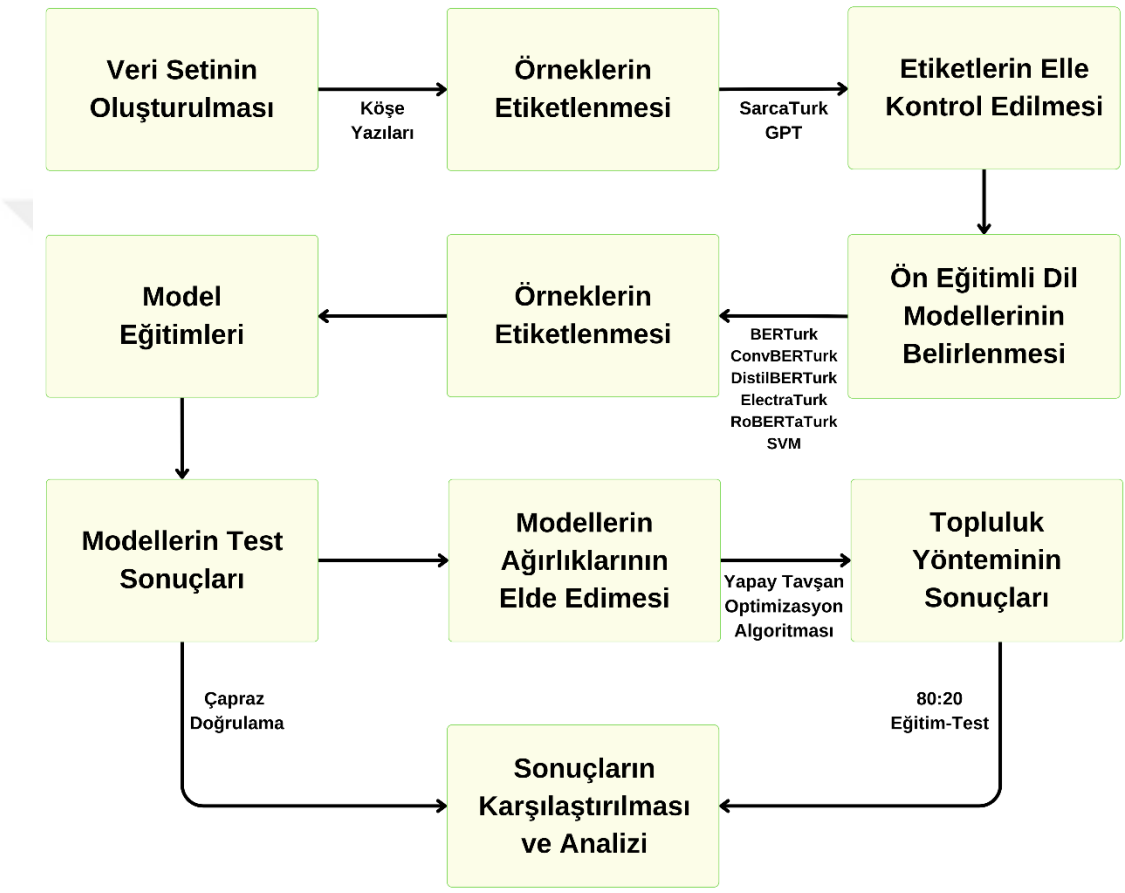
G. Li vd., 2022	İngilizce	RGAT	F1-skor: 72,25	IACV1
Chauhan vd., 2022	İngilizce	SEEmoji	F1-skor: 76,70	MUStARD
Cekinel ve Karagöz, 2022	İngilizce	BERT	Doğruluk: 78,60	Reddit
Savini ve Caragea, 2022	İngilizce	BERT + transfer öğrenme	F1-skor: 77,53	SARC
Misra ve Arora, 2023	İngilizce	CNN+LSTM	Doğruluk: 89,7	Kendi Veri Kümesi
Yao vd., 2023	İngilizce	M3N2	Doğruluk: 83,23	Hazır Veri Kümesi
Tasnia vd., 2023	İspanyolca	Hibrit	Doğruluk: 86,60	IberEval
Pandey ve Singh, 2023	İngilizce/Hintçe	BERT-LSTM	F1-skor: 95,50	Hazır Veri Kümesi
Yue vd., 2023	İngilizce	KnowleNet	Doğruluk: 88,87	Hazır Veri Kümesi
Grover ve Banati, 2024	İngilizce	Hibrit	F1-skor: 54,95	SarcOjiTest1
Helal vd., 2024	İngilizce	RoBERTa, DistilBERT+ bağlam	Doğruluk: 89,00	MUStARD
H. Liu vd., 2024	İngilizce	SAHFN	Doğruluk: 61,59	Multibully
Zhong vd., 2024	İngilizce	SEF	Doğruluk: 88,45	Hazır Veri Kümesi
Chakravarthi, 2025	Tamilce	TamBERT	Doğruluk: 79,80	Kendi Veri Kümesi
Q. Li vd., 2025	İngilizce	Sarcasm-GPT	Doğruluk: 82,31	MUStARD
Sukhavasi ve Dondeti, 2025	İngilizce	EnsembleNet	Doğruluk: 98,64	Memotion 7k
Y. Zhang vd., 2025	İngilizce	PIM-MS	Doğruluk: 64,96	Multibully
Taşlıoğlu ve Karagöz, 2017	Türkçe	Bilgi kazancı + RO	F1-skor: 88,60	Kendi Veri Kümesi
Dülger, 2018	Türkçe	Çok Katmanlı Algılayıcı	Doğruluk: 88,19	Hazır Veri Kümesi
Toçoğlu ve Onan, 2019	Türkçe	n-gram + TF + SVM	Doğruluk: 89,70	Kendi Veri Kümesi
Karabaş ve Diri, 2020	Türkçe	Attention BiLSTM	Doğruluk: 84,30	Kendi Veri Kümesi
Cemek vd., 2020	Türkçe	BERT	Doğruluk: 87,50	Kendi Veri Kümesi
Öztürk vd., 2021	Türkçe	BERT,LSTM	Doğruluk: 69,00	IronyTR

Çizelge 2.1. İğneleme ve ironi tespiti için gerçekleştirilen çalışmalarda kullanılan yöntemler, sonuçlar ve kullanılan veri kümeleri (devam)

Büyükbaş vd. 2021	Türkçe	SHAP, LIME, BERT	Doğruluk: 69,00	IronyTR
Onan ve Toçoğlu, 2022	Türkçe	RNN-AM	Doğruluk: 97,72	Kendi Veri Kümesi
Öztürk, 2024	Türkçe	BERT,SVM	Doğruluk: 69,00	IronyTR
Öztürk, 2024	Türkçe	BERT,SVM	Doğruluk: 99,00	Kaggle+Zaytung
Kara ve Karagöz, 2025	Türkçe		Doğruluk: 76,70	IronyTR

3. MATERYAL ve YÖNTEM

Günümüze kadar gerçekleştirilen mevcut çalışmaların rehberliğinde Türkçe ironi ve iğneleme içeren ifadelerin tespitine yönelik gerçekleştirilen tez çalışmasında kullanılan yöntem ve araçlar bu bölümde anlatılarak detaylandırılacaktır. Çalışma boyunca izlenen yöntem ve süreç Şekil 3.1'de sunulmuştur.



Şekil 3.1. Çalışmada izlenilen süreç ve çalışmanın akış şeması

Şekil 3.1'de görüleceği üzere tez çalışmasında uygulanan yöntem ve uygulama esnasında kullanılan araçlar, veri kümesinin oluşturulması, çalışmada kullanılan dil modelleri, yeni model olarak önerilen kolektif öğrenme için kullanılan optimizasyon algoritması, makine öğrenmesi modelinin, ön eğitimli dil modellerinin ve kolektif öğrenme yönteminin performans testlerinin gerçekleştirilmesi ve karşılaştırmalı analizi gerçekleştirilmiştir.

3.1. Veri Kümesi

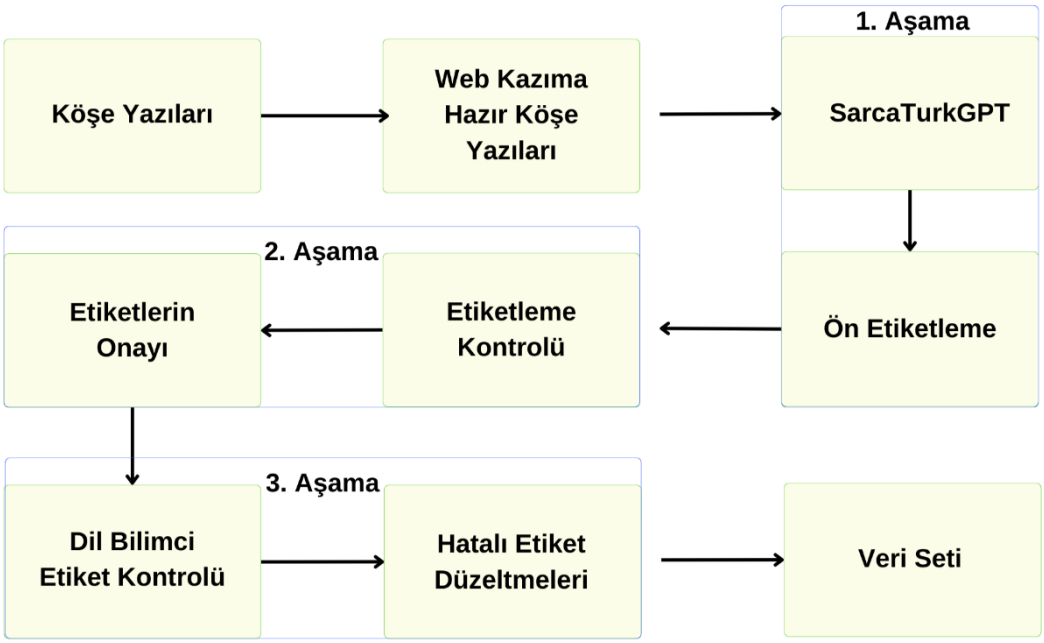
Günümüzde 200 milyonu aşkın insan tarafından konuşulan Türkçe, Balkanlardan Büyük Okyanus'a, Kuzey Buz Denizi'nden Tibet'e kadar geniş bir coğrafyada varlığını sürdürmektedir. Bu kadar geniş coğrafyada varlığının bulunmasının temel nedenlerinden biri tarihsel göçler sayesinde Avrupa, Amerika ve Avustralya kıtalarına kadar ulaşmış olmasıdır. Bu durum Türkçeyi dünyanın en geniş coğrafyaya yayılan dillerinden biri haline getirmiştir (Akalın, 2009; Özyetgin, 2006). Türkçe, yirmiden fazla yazı diliyle konuşulmakta olup, Türkiye Türkçesi, Azerbaycan Türkçesi, Türkmen Türkçesi, Özbek Türkçesi, Kazak Türkçesi, Kırgız Türkçesi ve Uygur Türkçesi gibi lehçeler arasında büyük ölçüde ortaklık bulunmaktadır (Akalın, 2009). Türkçe, köken bakımından genellikle Altay dilleri sınıfında bulunmakta ve Ural-Altay dil ailesiyle ilişkilendirilmiştir. Bu bağlamda Türkçe; Moğolca ve Mançu-Tunguzca ile yakın bir yapı gösterirken, bazı araştırmacılar Korece ve Japoncayı da bu aileye dâhil etmektedir (Ercalisun, 2004). Türkçe yazılı bir dil olarak dünya dilleri arasında yerini Orhun Yazıtlarına kadar uzanan köklü geçmişine dayanarak almaktadır. İleriki dönemlerde Kutadgu Bilig, Divan-ü Lügat-it Türk gibi eserler yine Türkçenin dünya dilleri arasında köklü bir geçmişe sahip olduğunu göstermektedir (Ercilasun, 2013; Özyetgin, 2006). Yapısal açıdan, sözcükler kök ve eklerin birleştirilmesi ile oluşması nedeniyle eklemeli bir dil olan Türkçe, cümle dizilişi özne-nesne-yüklem şeklindedir (Ercilasun, 2013). Kelime türetmede eklemenin yanı sıra birleştirme, kısaltma, ödünçleme ve tür değiştirme gibi yollar da kullanılmaktadır.

Türkçenin köklü geçmişi, geniş bir alanda konuşulması ve eklemeli yapısı, onu dünya dilleri arasında önemli bir konuma getirse de bu özellikler aynı zamanda doğal dil işleme açısından önemli zorluklara neden olmaktadır. Sözcüklerin köklere eklenen çok sayıda ek aracılığıyla türetilmesi, tek bir kökten birçok farklı kelimenin ortaya çıkmasına neden olur. Basit köke sahip bir kelimenin ekleme sonucunda uzun ve karmaşık kelimelere dönüşmesi bilgisayarların kelimeleri tanımasını ve çözümlemesini zorlaştırabilmektedir. Ayrıca Türkçede eş anlamlılık, mecaz, deyim ve bir kelimenin birden fazla anlama gelmesi gibi anlamsal açıdan çeşitliliklerin fazla olması bir kelimenin bağlama göre farklı anlamlar kazanmasına neden olur. Kültürel çeşitlilik nedeniyle Aynı kavram farklı bölgelerde farklı sözcüklerle ifade edilebilmesi yine

anlamın anlaşılmasında zorlayıcı olabilmektedir. Özellikle ironi ve iğneleme gibi sözün gerçek anlamından farklı bir niyet taşıyan söylem biçimleri Türkçenin anlamsal açıdan zorlayıcılığını göstermektedir. Bu tür ifadelerde kelimeler basit anlamda olumlu görünse de bağlam, tonlama, bulunulan konum, zaman dikkate alındığında farklı bir anlam kazanabilir. Zaman zaman insan için bile ironi veya iğnelemenin anlaşılması zorken hesaplamalı yöntemlerin bu anlamsal farklılıkları yakalaması oldukça zordur.

3.1.1. Veri kümesinin oluşturulması

Tez çalışmasında Türkçe metinlerde ironi ve iğneleyici ifadelerin tespiti için köşe yazıları kaynak alınarak veri kümesi oluşturulmuştur. 3 sınıflı veri kümesinin oluşturulmasına yönelik akış şeması Şekil 3.2'de sunulmuştur.



Şekil 3.2. Veri kümesi oluşturma işleminin akış şeması

Veri kümesi elde edilmesi amacıyla;

- http://www.kemik.yildiz.edu.tr adresinden elde edilen 50 yazara ait 50 köşe yazısı öncelikle ele alınmıştır.
- İkinci olarak gazete internet sayfaları üzerinde web kazıma işlemi gerçekleştirilmiştir.

- Web kazıma yöntemleri Visual Studio Code üzerinde gerçekleştirilmiştir.
- Web kazıma işlemlerinde BeautifulSoup (Richardson, 2025) kütüphanesi kullanılmıştır.
- Yapay zekâ asistanı olarak SarcaTurkGPT (Duran, 2025) kullanılmıştır.
- Örneklerin etiketleme işlemi dil bilimci tarafından kontrol edilmiştir.

3.1.2. Veri kümesinin etiketlenmesi

Veri kümesinin etiketlenmesi, yapay zekâ asistanının ön etiketleme yapması, insan kullanıcının etiketlemeleri kontrol ederek eleme işleminden geçirmesi, dil bilimci yardımıyla etiketlemesi yapılan örneklerin uzman tarafından işlenmesi ve varsa yanlış etiketlerin düzeltilmesi olmak üzere 4 aşamada gerçekleşmektedir.

1. Aşama yapay zekâ asistanı ile ön etiketleme

Geleneksel veri etiketlemenin zaman ve kaynaklar açısından maliyetli olması nedeniyle, yapay zekâ destekli etiketlemelerin varlığı ön plana çıkmaktadır. Etiketleme türleri olarak sıfır örnekli etiketleme, az örnekli etiketleme, talimat tabanlı etiketleme gibi örnekler gösterilebilir. Talimat tabanlı etiketleme literatürde talimat tabanlı ince ayar veya denetimli ince ayar olarak da isimlendirilmektedir. Bu ince ayar sonraki kelimeyi tahmin etme veya bilgi üretme amacıyla eğitilen büyük dil modellerinin, kullanıcılar tarafından belirli talimatları takip ederek örnek etiketlemesinde kullanılmaktadır. Yöntemde talimat ve çıktı bölümlerinden oluşur ve belirli örnekler üzerinden denetimli ince ayar yapılmasına dayanır. Böylelikle ince ayar yapılan dil modelinin talimatlar doğrultusunda belirli sınırlamalarla istenilen çıktının üretilmesi sağlanmaktadır (S. Zhang vd., 2023). İncelenmesi ve analiz edilmesi gereken metinlerin büyüklüğü nedeniyle, talimat tabanlı etiketleme yöntemi tercih edilmiştir. Dil bilimci tarafından ironi ve iğneleme içeren ifadelerin tespiti için büyük dil modeli (OpenAI GPT-4o) kullanılarak oluşturulan SarcaTurkGPT isimli bir asistan kullanılmıştır. Asistan kendisine verilen köşe yazısı metinlerini analiz ederek içerisinde ironi ve iğneleme olabilecek metinleri işaret ederek ön etiket yapılmasını sağlamaktadır.

2. Aşama kullanıcı etiketlemeleri

SarcaTurkGPT tarafından ön etiketlemesi yapılan örnekler insan kullanıcı tarafından kontrol edilerek uygun sınıfa ait olarak işaretlenmektedir.

3. Aşama dil bilimci etiket kontrolü

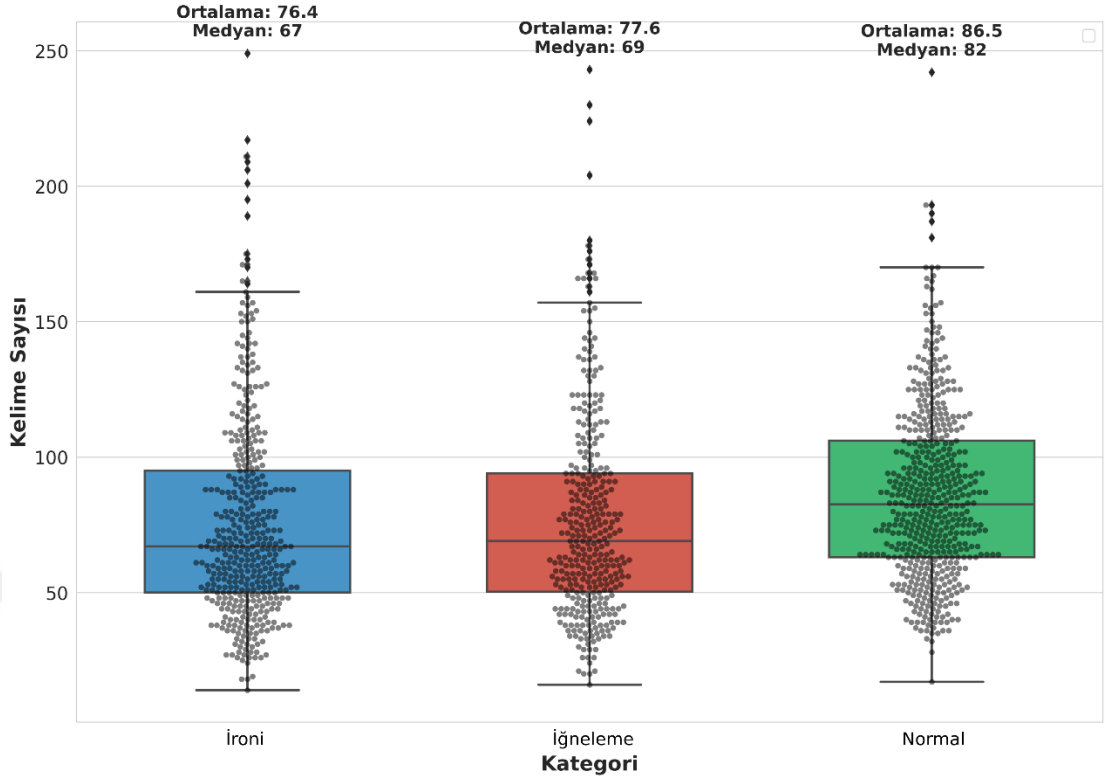
SarcaTurkGPT ve insan kullanıcıların etiketleme sürecinden sonra dil bilimci tarafından örneklerin kontrolü sağlanmaktadır. Böylelikle etiketleme süreci yarı-denetimli hale gelmektedir.

4. Aşama kullanıcı etiketlemeleri

Dil bilimci tarafından kontrolü sağlanan örneklerin etiketlemesinde hata olması durumunda insan kullanıcı tarafından tekrar incelenmesi ve etiketlemesi dil bilimcinin önerilerine göre gerçekleştirilmektedir.

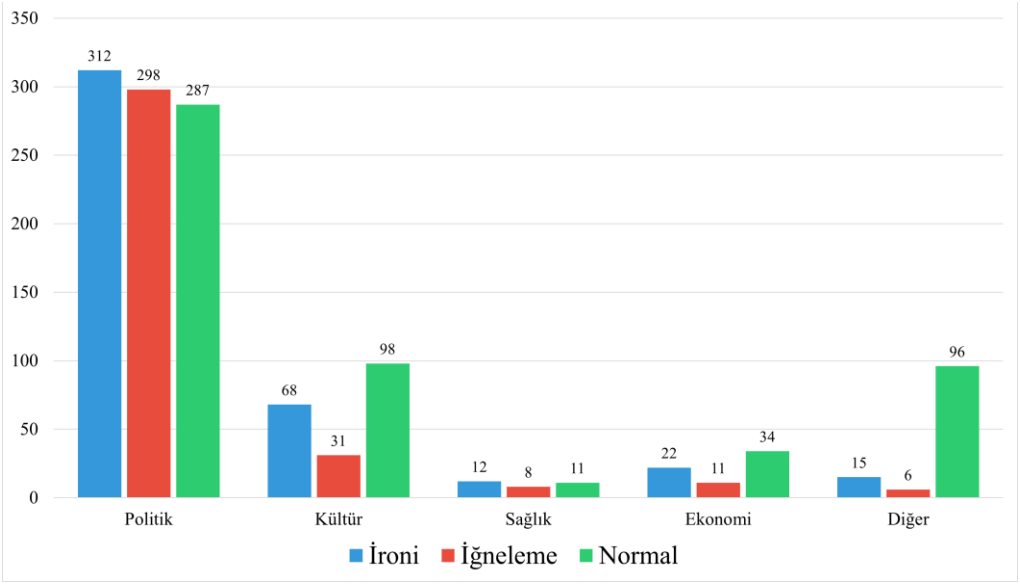
4 aşamalı etiketleme sürecinin ardından 429 örnek ironi, 354 örnek iğneleme ve 526 örnek normal sınıflarına ait veri kümesi oluşturulmuştur.

Veri kümesinin oluşturulması sırasında sınıflara ait örneklerin dağılımları, kelime sayılarının ortalaması, minimum ve maksimum kelime sayıları gibi modelin öğrenme sırasında performansını etkileyen istatistiksel dağılımlara dikkat edilmiştir. Çalışmada kullanılan veri kümesinin sınıflara göre temel istatistikleri Şekil 3.3'te sunulmuştur.



Şekil 3.3. Sınıflara göre temel istatistikler

Veri kümesinin oluşturulması sırasında sınıflara ait örneklerin politik, kültür, sağlık, ekonomi ve diğer kategoriler arasındaki istatistiksel dağılımları Şekil 3.4'te sunulmuştur.



Şekil 3.4. Veri kümesinde kullanılan metinlerin kategorileri

3.1.3. Alt veri kümelerinin oluşturulması

Türkçede ironik veya iğneleyici ifadelerin tespitinde konuşmanın geçmişi yani bağlamı son derece önemlidir. İroni ve iğneleme içeren mevcut cümle zaman zaman bu ifadelerin anlaşılması için yeterli olmayabilir. Bu nedenle bu cümleden önceki cümlelerde ironik veya iğneleyici ifadenin hazırlığı olabilir. Geçmiş bağlamın ironik ve iğneleyici ifadelerin tespit edilmesindeki öneminin araştırılması amacıyla mevcut veri kümesinden yola çıkarak bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümeleri oluşturulmuştur. bir-gram veri kümesi yalnızca ironik, iğneleyici veya normal cümleleri içermektedir. iki-gram veri kümesinde, ironik, iğneleyici veya normal cümlenin yanı sıra bir önceki cümle de bulunmaktadır. üç-gram, dört-gram ve tüm-gram veri kümeleri ise sırasıyla ironik, iğneleyici veya normal cümleden önce gelen iki, üç ve tüm cümlelerin dahil edilmesiyle oluşturulmuştur. Bu yapılandırma sayesinde, ironik veya iğneleyici cümleden önce gelen cümlelerin hedef sınıf üzerindeki etkisi araştırılmıştır. bir-gram, iki-gram ve üç-gram veri kümelerinin örnek mimarisi Çizelge 3.1'de sunulmuştur.

Çizelge 3.1. Sınıflara göre veri kümesi mimarileri

Sınıf	bir-gram	iki-gram	üç-gram	Paragraf
İroni	Böylece siz de bir Danimarkalı kadar mutlu olabilirsiniz.	Yumuşak bir aydınlatma, hoş kokulu mumlar, lezzetli yemekler, samimi sohbetler, sıcak tutan yumuşak giysiler… Böylece siz de bir Danimarkalı kadar mutlu olabilirsiniz.	Üstelik yapılması gerekenler hiç de zor sayılmazdı. Yumuşak bir aydınlatma, hoş kokulu mumlar, lezzetli yemekler, samimi sohbetler, sıcak tutan yumuşak giysiler… Böylece siz de bir Danimarkalı kadar mutlu olabilirsiniz.	Kitabı çok satanlar listesine taşıyan ve Türkçe dahil pek çok dilde yayınlanıp popülerleşmesini sağlayan da kuşkusuz insanların bu mutluluk formüllerini öğrenme ihtiyacıydı. Üstelik yapılması gerekenler hiç de zor sayılmazdı. Yumuşak bir aydınlatma, hoş kokulu mumlar, lezzetli yemekler, samimi sohbetler, sıcak tutan yumuşak giysiler… Böylece siz de bir Danimarkalı kadar mutlu olabilirsiniz.
İğneleme	Ne de olsa “Alamancı” onun bakışı bir başka...	Bir de dizileri uzun tutuyorlar, kısa yapsınlar! diyordu, o sevimli Türkçesiyle. Ne de olsa “Alamancı” onun bakışı bir başka...	Üç yıldır Türkiye’deyim, ilk defa bir ödül alıyorum. Bir de dizileri uzun tutuyorlar, kısa yapsınlar! diyordu, o sevimli Türkçesiyle. Ne de olsa “Alamancı” onun bakışı bir başka...	Oysa “Hürrem Sultan” Meryem Uzerli’nin dediği başka. Üç yıldır Türkiye’deyim, ilk defa bir ödül alıyorum. Bir de dizileri uzun tutuyorlar, kısa yapsınlar! diyordu, o sevimli Türkçesiyle. Ne de olsa “Alamancı” onun bakışı bir başka...
Normal	Kanalların ve yönetimlerinin geleceğini belirliyor, yapım şirketlerinin milyonlarca dolar kâr veya zarar etmesine neden oluyor.	Her bir rating altın değil pırlanta değerinde. Kanalların ve yönetimlerinin geleceğini belirliyor, yapım şirketlerinin milyonlarca dolar kâr veya zarar etmesine neden oluyor.	Her gün rating karnesi alan onlarca yapım şirketi kıyasıya mücadele ediyorlar. Her bir rating altın değil pırlanta değerinde. Kanalların ve yönetimlerinin geleceğini belirliyor, yapım şirketlerinin milyonlarca dolar kâr veya zarar etmesine neden oluyor.	Şaşkınlık verici çünkü Türkiye’de rekabetin en sert ve en görünür olduğu yerlerden biri dizi sektörü. Her gün rating karnesi alan onlarca yapım şirketi kıyasıya mücadele ediyorlar. Her bir rating altın değil pırlanta değerinde. Kanalların ve yönetimlerinin geleceğini belirliyor, yapım şirketlerinin milyonlarca dolar kâr veya zarar etmesine neden oluyor.

Çizelge 3.1 incelendiğinde bir-gram bir cümlelik, iki-gram iki, üç-gram üç, dört-gram dört cümlelik ve tüm-gram için tüm paragrafı kapsayan örnekler içermektedir. Bu mimariyi kullanarak deneysel çalışmalar gerçekleştirilmiştir.

3.1.4. Eğitim ve test veri kümelerinin hazırlanması

Makine öğrenmesi uygulamalarında modelin güvenilir şekilde eğitilebilmesi için kullanılan veri kümesinin belirli yöntemlerle en az iki bölüme ayrılması gerekir. Eğitim kümesi, modelin öğrenme sürecinde kullandığı verilerden oluşur. Test kümesi, modelin daha önce hiç görmediği örnekler üzerinde test edilerek genelleme yeteneğini ölçer ve eğitim sırasında kullanılmaz. Bazı durumlarda ayrıca bir doğrulama kümesi de ayrılır; bu küme, modelin hiper parametrelerinin ayarlanmasında ve aşırı öğrenmenin önlenmesinde önemlidir. Veri kümesinin bölünmesi sayesinde modelin hem öğrenme başarısı hem de gerçek hayattaki performansı daha sağlıklı bir şekilde ölçülebilir (Birba, 2020). Çalışmada veri kümesinin bölünmesinde aşağıda belirtilen iki yöntem seçilmiştir.

Oransal Bölünme

Modellerin eğitiminde kullanılacak olan veri kümesinin bir kısmının eğitim ve kalan kısmının test olarak kullanıldığı bölünme türüdür. Modelin genelleme yeteneği açısından eğitim veri kümesinin örnek sayısı önemlidir. Makine öğrenmesi modellerinde verilerin bölünme oranı genellikle 80:20 olarak belirlenmektedir (Bouktif vd., 2018). Çalışmada 80:20 oranında eğitim-test bölünmesi tercih edilerek model eğitimleri gerçekleştirilip, test edilmiştir. Bölünme sırasında eğitim ve test veri kümelerinde sınıf dağılımları dengeli olarak tutulmuştur.

k-kat Çapraz Doğrulama

k-katlı çapraz doğrulama, verinin k parçaya ayrılmasıyla uygulanır. Modeller k-1 parça üzerinde eğitilir, kalan parça üzerinde test edilir ve bu işlem tüm parçalar için tekrarlanır. Böylece tüm veri üzerinde performans ölçümü yapılmış olur. Bu yöntem, aşırı öğrenmeyi önlemek ve modelin yeni verilerdeki başarısını artırmak için kullanılır.

Model seçiminde ve hiper parametre ayarlamada avantaj sağlarken, k değeri çok yüksek seçildiğinde gereksiz hesaplama maliyetine yol açabilir (Teodorescu ve Obreja Brasoveanu, 2025). Model sonuçlarının değerlendirilmesinde güvenilir ve tutarlı yapılmasını sağlamak için literatürde kat değeri sıklıkla kullanılan 10 olarak uygulanmıştır.

3.2. Modelleme

Türkçe köşe yazılarında ironik ve iğneleyici ifadelerin tespit edilmesi için, ön işlemden geçirilmiş ve farklı mimarilerde hazırlanmış bağlama duyarlı veri kümeleri için uygun modellerin belirlenmesine ve çalışmada önerilen kolektif temelli yaklaşımın hazırlanmasına bu bölümde yer verilmektedir.

Türkçe ve diğer diller için yapılan mevcut çalışmalar incelendiğinde birçok farklı kelime gömme yönteminin, klasik makine öğrenmesi yöntemlerinin, derin öğrenme yöntemlerinin, önceden eğitilmiş dil modellerinin, farklı mimarilerin ve yaklaşımların kullanıldığı hibrit yöntemlerin, başarıyı artırmak için kolektif öğrenme yöntemlerinin kullanıldığı gözlemlenmiştir. Çalışmalarda kullanılan yöntemler Çizelge 3.2'de sunulmuştur.

Çizelge 3.2. Yapılan çalışmalarda kullanılan yöntemlerin kategorilere göre sınıflandırılması

Kategori	**Yöntemler**
Kelime gömme yöntemleri	GloVe, ELMo, BERT, Flair, Word2Vec, FastText
Makine öğrenmesi yöntemleri	Destek Vektör Makineleri, Rastgele Orman, Naive Bayes, Regresyon Ağaçları, Karar Ağaçları, Tekrarlanan Aşamalı Budama, Lojistik Regresyon, Yapay Sinir Ağları, Lojistik Model Ağaçları, k-En Yakın Komşu
Derin öğrenme yöntemleri	Evrişimli Sinir Ağları, Tekrarlayan Sinir Ağları, Uzun Kısa Süreli Bellek, Çift Yönlü Uzun Kısa Süreli Bellek, Kapılı Tekrarlayan Birim, Çift Yönlü Kapılı Tekrarlayan Birim
Önceden eğitilmiş yöntemler	BERT, ConvBERT, RoBERTa, DistilBERT, ELECTRA
Hibrit yöntemler	BiLSTM-CNN, RCNN-RoBERTa, CNN-LSTM, BERT-Transfer Öğrenme, BERT-LSTM

Yapılan çalışmaların incelenmesi sonucunda ele alınan ironi ve iğneleme problemine yaklaşımın sonuçlar üzerinde önemli etkisi olduğu görülmektedir. Yaklaşım olarak ikili sınıflandırma seçildiğinde performansın arttığı, nadir de olsa ironi ve iğneleme sınıflarının aynı anda tespitinde performansın düştüğü görülmektedir. Diğer bir durum ise kullanılan veri kümesinin zorluğu performans açısından önemli olduğu, yapılan çalışmalarda aynı yöntemin farklı veri kümeleri üzerinde birbirinden uzak performans metrikleri elde ettiği görülmektedir. Kullanılan yöntemlerinde belirtilen durumlara göre performanslarının farklılaştığı, önceden eğitilmiş dil modellerinin, makine öğrenmesi yöntemlerinin, derin öğrenme yöntemlerinin ve hibrit yöntemlerin farklı veri kümeleri üzerinde performanslarının değiştiği görülmektedir. Ancak yine de önceden eğitilmiş dil modellerinin ince ayar yapılarak kullanıldığı çalışmalar son yıllarda önem kazanmıştır.

3.2.1. Destek vektör makineleri

Destek vektör makineleri (Support Vector Machines-SVM), sınırlı sayıda öğrenme örüntüsü üzerinden iyi bir genelleme düzeyi elde etmek amacıyla Yapısal Risk Minimizasyonu (YRM) tümevarım prensibini uygulayan bir öğrenme makinesidir. SVM, karmaşık veri kümelerinde, çözümlemesi zor örüntülerin tanımlanmasında kullanışlı bir öğrenme algoritmasını uygulamaktadır (Tasar vd., 2019). SVM'nin ilgi çekici özelliği; veri kümesi üzerinde ortalama hata karesini minimize ederek türetilen ampirik risk minimizasyonu prensibinden ziyade, istatistiksel öğrenme teorisindeki yapısal risk minimizasyonu niteliğinde çalışmasıdır. SVM sınıflandırma ve regresyon gibi farklı problemler üzerinde çalıştırılabilir (Yakut vd., 2014). SVM doğrusal olmayan haritalama fonksiyonları yardımıyla orijinal girdi uzayından yüksek boyutlu nitelik uzayına dönüşüm yapar. Bu yeni uzayda veriler doğrusal olarak sınıflandırılabilir (Yabanova ve Yumurtacı, 2018). Bu dönüşüm sürecinde çekirdek fonksiyonları kullanılır. Fonksiyonlar, yüksek boyutlu nitelik uzayı ile uğraşmadan değerlerin tekrarlı şekilde hesaplanması yerine çekirdek fonksiyonunda değerin yerine yerleştirilmesiyle nitelik uzayındaki değerin bulunmasını sağlar (Yakut vd., 2014).

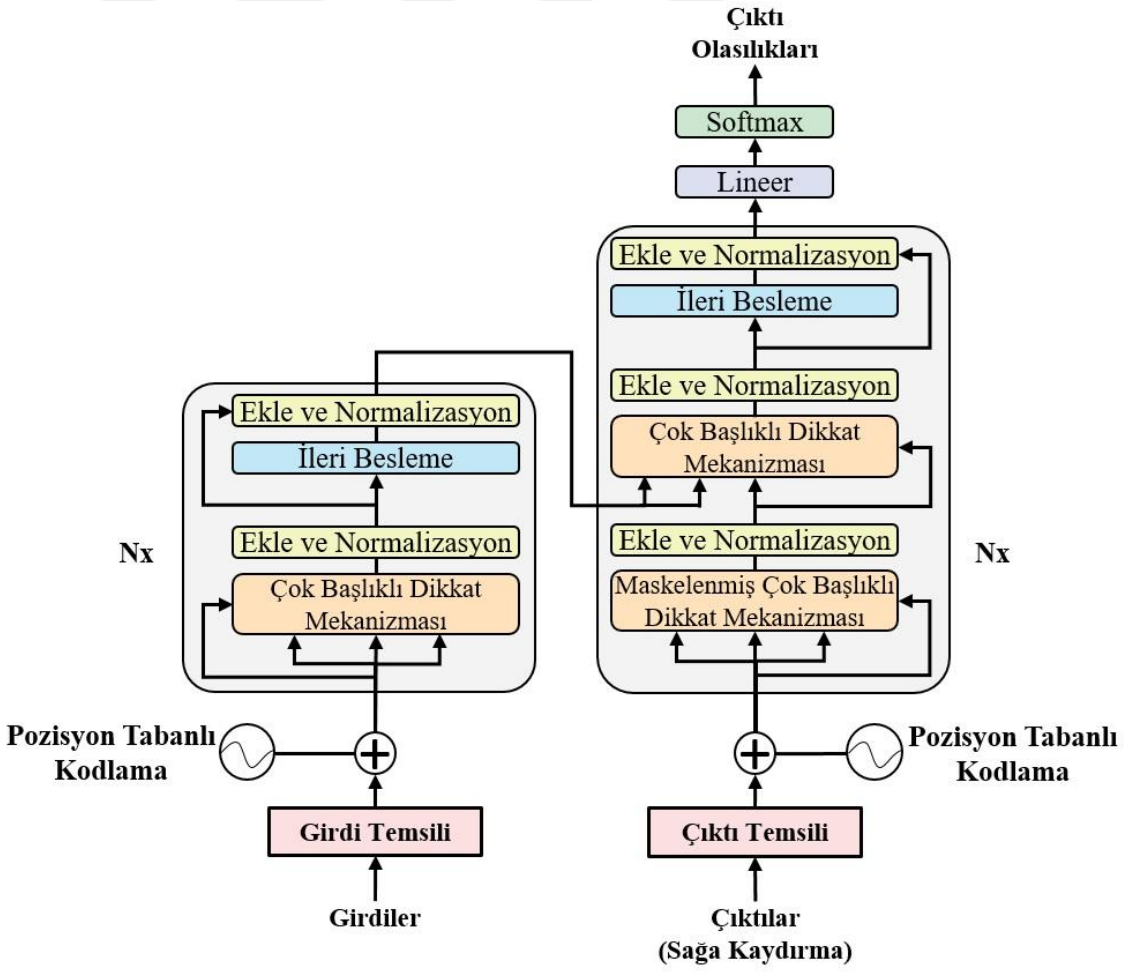
SVM'de genellikle dört çekirdek fonksiyon kullanılır. Bunlar:

- Doğrusal
- Polinominal
- Sigmoid
- Radyal Tabanlı Fonksiyon

3.2.2. BERTurk

Çift yönlü kodlayıcı gösterimleri (BERT), Devlin ve arkadaşları, sol ve sağ bağlamı aynı anda dikkate alarak dil temsillerini öğrenebilen ve doğal dil işleme alanında devrim yaratan bir modeldir (Devlin vd., 2019). Temel özelliklerinden biri, kelimeleri iki yönlü olarak anlamasıdır. Bu sayede bağlamı daha iyi kavrayarak anlam aktarımında etkili sonuçlar üretir. Önceki tek yönlü modellerde mümkün olmayan bu özellik dilin bağlamsal özelliklerini daha iyi yakalayarak Model, metin sınıflandırma, duygu analizi, soru-cevap sistemleri gibi çeşitli görevlerde yüksek başarı oranları gösterir. BERT,

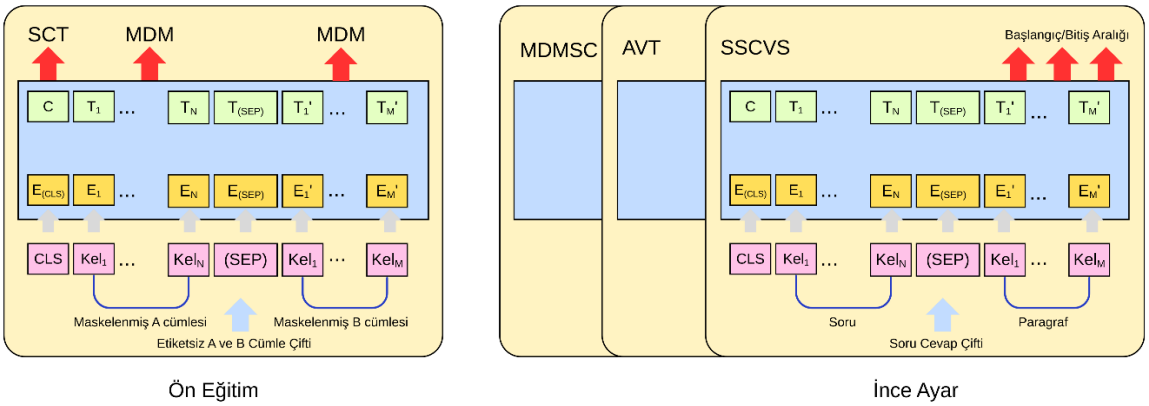
etiketlenmemiş metinleri kullanarak her iki yönden bağlamı dikkate alan temsiller üretmekte ve bir ek katmanla farklı sınıflandırma, soru-cevap gibi farklı görevler için yeniden kullanılabilmektedir. Modelin temelini, Vaswani ve arkadaşları (Vaswani vd., 2017) tarafından geliştirilen dönüştürücü mimarisi, tekrar eden veya evrişimsel yapılar yerine çok başlıklı öz-dikkat mekanizması kullanarak metinler gibi sıralı verilerde bulunan uzun mesafeli bağımlılıkların öğrenilmesini sağlar. Kullanılan dikkat mekanizması girdideki kelimenin diğer kelimelerle olan yakınlığını öğrenmesini ve buna göre hangi kelimelere daha fazla odaklanacağını belirler. Bu ağırlıklandırmayı bir sorguyu anahtar-değer çiftleriyle ilişkilendirip çıktıyı ağırlıklı toplam şeklinde üreterek gerçekleştirir. Bu yaklaşım, bağımlılıkların daha kolay öğrenilmesini sağlarken, metinlerin işlenmesini paralel hale getirerek eğitim süresini kısaltmaktadır. Dönüştürücü mimarisi ve içerisinde bulunan katmanlar Şekil 3.5'te sunulmuştur.



Şekil 3.5. Dönüştürücü mimarisi (Çağla, 2024)

Dönüştürücü mimarisi encoder ve decoder bölümlerinden oluşurken, kodlayıcı katmanında, çok başlıklı dikkat mekanizması her kelimenin cümledeki diğer kelimelerle ilişkisini kurar, ardından basit ileri beslemeli ağ bu bilgiyi işler. Çözücü katmanı da benzer bir yapıya sahip olmakla birlikte encoderdan gelen bilginin girdi olarak alındığı dikkat katmanı bulunur. Dönüştürücü mimarisi farklı uygulamalarda kodlayıcı olarak, çözücü olarak ve her ikisini birlikte kullanılmakta olup, metin sınıflandırma problemlerinde sadece kodlayıcı kısmı kullanılmaktadır (Çağla, 2024).

BERT modelinin eğitim süreci iki aşamalı olarak ele alınmaktadır. Öncelikle büyük ve etiketlenmemiş veriler üzerinde maskeli dil modelleme ve sonraki cümle tahmini görevleriyle ön-eğitim yapılır. Maskeli dil modellemesinde giriş değerlerinin bir yüzdelik kısmı rastgele maskelenir ve ardından maskeli değerler tahmin edilir. Bu işlemde model kelimeler arasındaki ilişkiyi anlar. Sonraki cümle tahmininde model, iki cümlenin birbirlerinin devamı olup olmadığını, ikinci cümlenin birincinin doğal devamı olduğu ve başka bir durumda rastgele bir cümle seçilerek devamı olmadığı şeklinde etiketlenir. Bu süreçte [CLS] belirteci sınıflandırma için, [SEP] belirteci ise iki cümleyi ayırmak için kullanılır. Böylece model, cümleler arası ilişkiyi öğrenir ve cümle düzeyinde temsil yeteneği kazanır (Çağla, 2024). İkinci aşamada ise model, görev özelinde etiketlenmiş veriyle yeniden eğitilerek ince ayar sürecinden geçirilir. İnce ayar aşamasında öğrenilen parametreler görev özelinde etiketli veriyle güncellenerek farklı bir göreve uyarlanabilir ve yüksek başarı sağlayabilir (Koroteev, 2021). Bu yaklaşım, modelin ürettiği bağlamsal gömmelerin kelimelerin kullanım bağlamına göre anlamlarını daha iyi yansıtmaktadır. Böylece duygu analizi, soru-cevaplama ya da sahte yorum tespiti gibi görevlerde doğal dil işleme uygulamalarına kolayca uyarlanabilmesini sağlamaktadır (Gardazi vd., 2025). BERT modelinin ön eğitim ve ince ayar aşamaları Şekil 3.6'da sunulmuştur. BERT modelinin $BERT_{TEMEL}$ ve $BERT_{BÜYÜK}$ olmak üzere iki versiyonu bulunmaktadır. Temel modelde 12 katman ve 110 milyon parametre bulunurken, büyük modelde 24 katman ve 340 milyon parametre bulunmaktadır (Devlin vd., 2019).

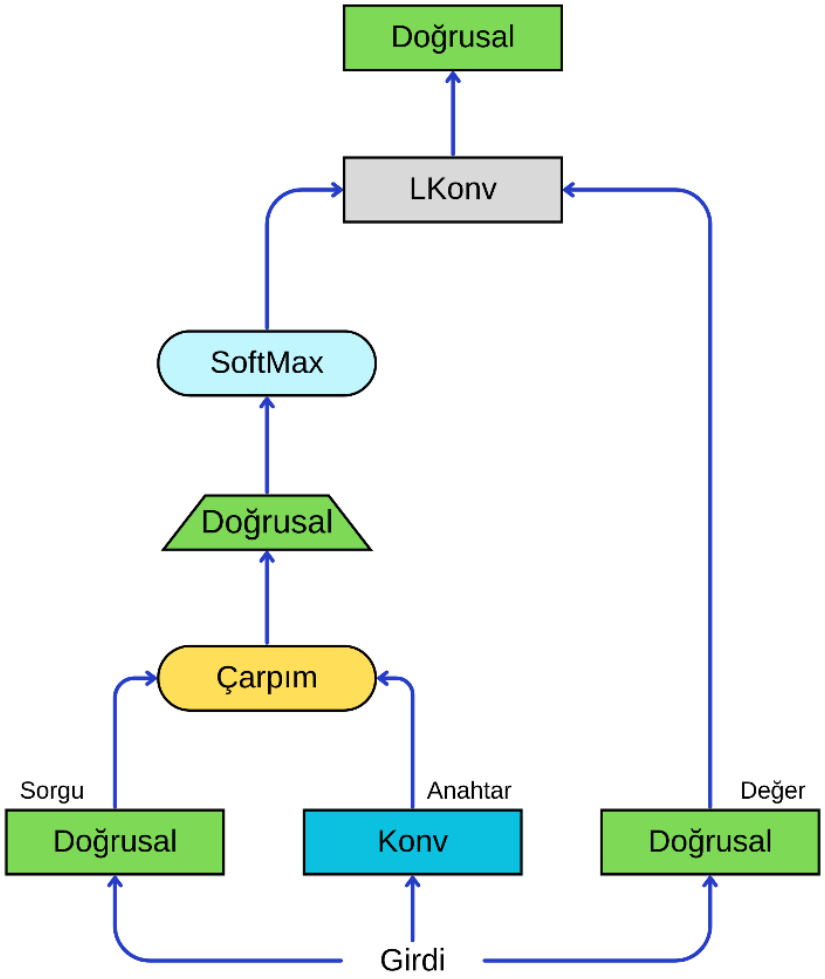


Şekil 3.6. BERT modelinin ön eğitim ve ince ayar aşamaları (Devlin vd., 2019'den değiştirilerek alınmıştır)

Farklı görevler için kolayca uyarlanabilmesi sebebiyle, Türkçenin dil yapısını ve özelliklerini güçlü bir şekilde analiz edebilmek için BERT mimarisi ve Türkçe veri kümesi kullanılarak BERTurk modeli eğitilmiştir. Bu bağlamda BERTurk, dilin morfolojik ve sentaktik özelliklerine uygun şekilde eğitilmiştir. 35 GB büyüklüğünde ve yaklaşık 4 milyar token içeren geniş bir Türkçe metin derlemiyle eğitilen model, veri kümelerinin çeşitliliği sayesinde farklı dil kullanımlarını ve bağlamları başarıyla anlamaktadır (Schweter, 2020).

3.2.3. ConvBERTurk

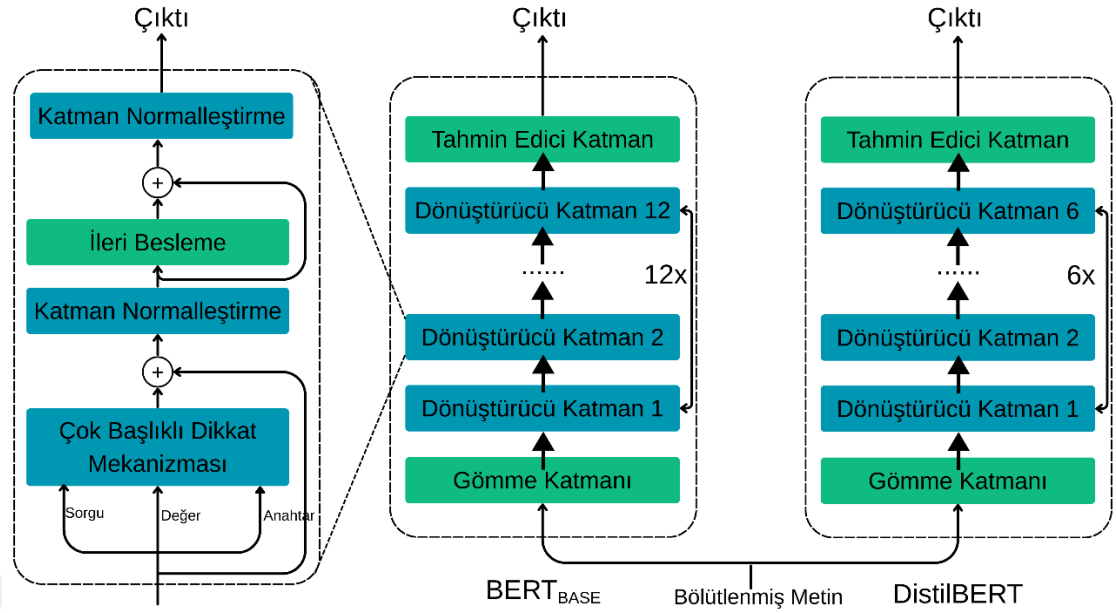
ConvBERTurk, doğal dil işleme alanında, BERT mimarisinin konvansiyonel katmanlarla birleştirilmesiyle geliştirilmiş bir dil modelidir (Jiang vd., 2020). Bu model, bağlamsal ve yerel özellikleri bir arada kullanarak daha anlamlı ve etkili dil işleme yetenekleri sunmayı amaçlar. ConvBERTurk modelinin yapısı Şekil 3.7'de sunulmuştur. ConvBERTurk ise Türkçe doğal dil işleme (NLP) için geliştirilmiş, Türkçeyi anlama ve işleme yeteneklerine sahip bir modeldir (dbmdz, 2020a). ConvBERT mimarisine dayanan bu model, 35 GB büyüklüğünde ve yaklaşık 4 milyar token içeren geniş bir Türkçe metin derlemiyle eğitilmiş; böylece dilin morfolojik ve sentaktik özelliklerine uygun hale getirilmiştir.



Şekil 3.7. ConvBERT mimari diyagramı (Jiang vd., 2020'den değiştirilerek alınmıştır)

3.2.4. DistilBERTurk

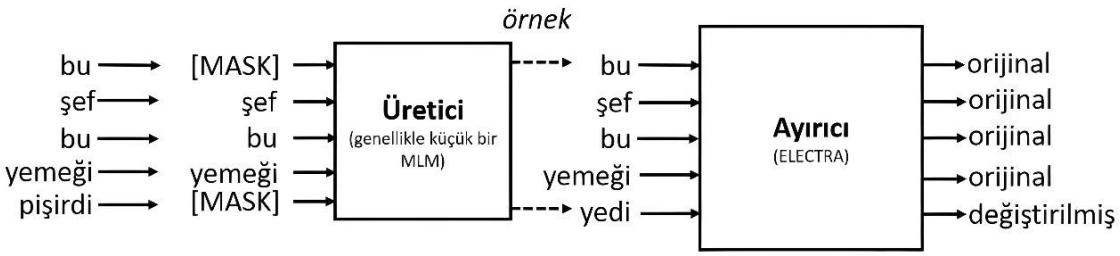
DistilBERT, BERT modelinin hafifletilmiş bir versiyonudur (Sanh vd., 2019). "Bilgi damıtma" teknikleriyle eğitilen bu model, daha hızlı çalışır ve daha az kaynak tüketerek doğal dil işleme (NLP) uygulamalarında verimlilik sağlar. DistilBERT, BERT'in temel yapı taşlarını korurken parametre sayısını azaltarak performans ve hız açısından önemli avantajlar sunar. Eğitim sürecinde, BERT'in büyük veri kümelerinden faydalanır ve "öğretmen" konumundaki BERT modelinden öğrenerek temel yeteneklerini kaybetmeden daha hafif bir yapı geliştirir (Abadeer, 2020). Modelin mimari yapısı Şekil 3.8'de sunulmuştur. Bu çalışmada, Türkçe veri kümesiyle eğitilmiş DistilBERTurk modeli kullanılmıştır (dbmdz, 2020b).



Şekil 3.8. DistilBERT modelinin mimarisi (J. Li vd., 2024'ten değiştirilerek alınmıştır)

3.2.5. ELECTRATurk

Token değişimlerini doğru bir şekilde sınıflandırmayı verimli biçimde öğrenen bir kodlayıcı (ELECTRA), 2020 yılında geliştirilmiş, önceden eğitilmiş bir dil modelidir (Clark vd., 2020). ELECTRA, BERT gibi modellerin hızı ve verimliliğini, yüksek performansla birleştirmeyi hedefler ve bu doğrultuda çok sayıda NLP görevinde üstün başarı sergiler (Husein vd., 2023). Modelin temelinde iki ana bileşen bulunur: üretici ve ayırıcı. Üretici, belirli kelimeleri değiştirmek için kullanılırken, ayırıcı bu değiştirilmiş kelimeleri analiz ederek doğru tahminlerde bulunmaya çalışır. ELECTRA modelinin kullandığı token yer değiştirmesi yöntemi ait Şekil 3.9'da sunulmuştur.



Şekil 3.9. ELECTRA modelinin token yer değiştirmesi (Çağla, 2024)

Bu yapı sayesinde ELECTRA, maskeli dil modelleme (MLM) yerine değiştirilmiş kelime tespiti (RTD) adı verilen bir görev üzerinde çalışır. Bu süreç, dil modelinin eğitim verimliliğini artırır ve özellikle söz dizimsel bağlamların öğrenilmesinde avantaj sağlar (Xu vd., 2020). ELECTRA'nın eğitim sürecinde, model yalnızca verileri değil, aynı zamanda verilerin bağlamını anlamakta ve doğru kelime tahminleri yapmaktadır. Yapılan araştırmalar, ELECTRA'nın geleneksel BERT eğitim yöntemlerinden belirgin bir avantaja sahip olduğunu ve daha az hesaplama gücü ile daha fazla bilgi öğrendiğini göstermektedir (Gargiulo vd., 2022). Türkçe veri kümesiyle eğitilerek Türkçe doğal dil işleme için oluşturulan ELECTRATurk modeli, bu çalışmada kullanılmıştır (dbmdz, 2020c).

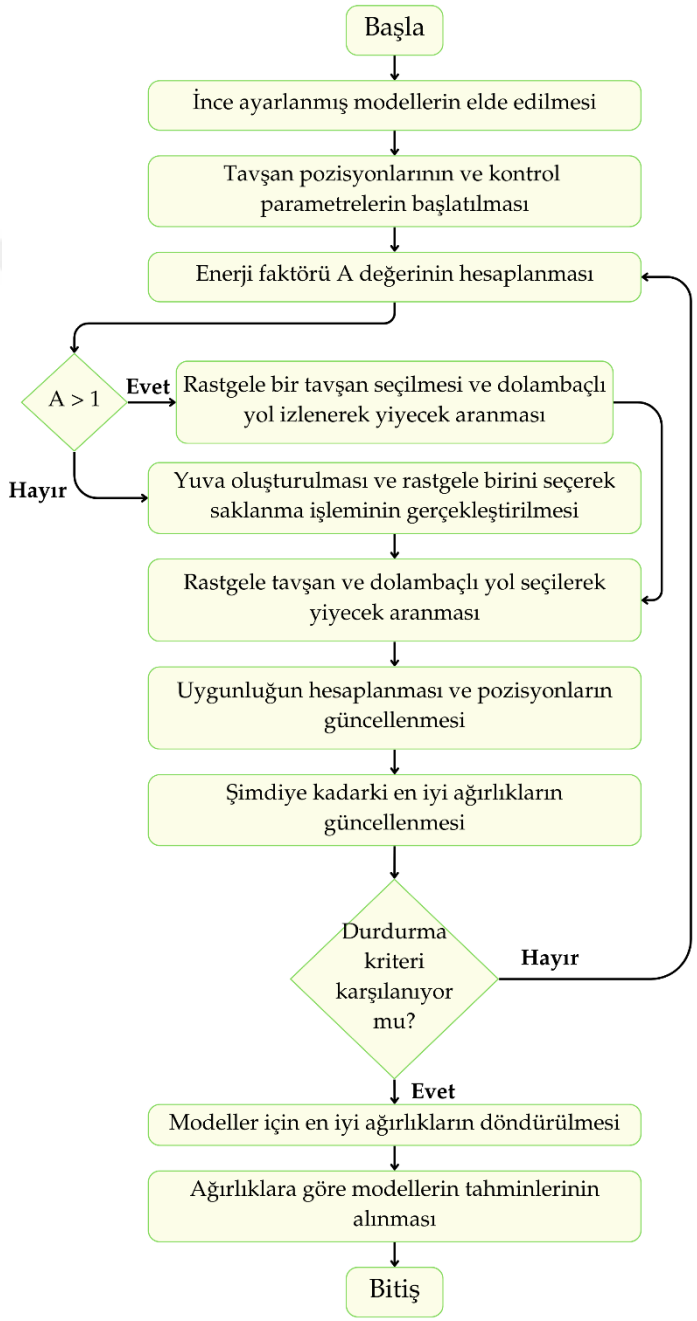
3.2.6. RoBERTaTurk

Önceden eğitilmiş, güçlendirilmiş ve optimize edilmiş BERT yaklaşımı (RoBERTa), BERT modelinin ön eğitim prosedürünü geliştirerek doğal dil işleme (NLP) alanında önemli bir başarı göstermiş, gelişmiş bir dönüştürücü tabanlı çift yönlü eğitim mekanizmasıdır. Bu model; daha uzun süreli ve daha büyük veri kümeleriyle eğitim, sonraki cümle tahmini (NSP) hedefinin kaldırılması, daha uzun dizilerle çalışma ve dinamik maskeleme yönteminin kullanılması gibi iyileştirmeler sunar. RoBERTa, BERT mimarisi ve eğitim yönteminde sistematik değişiklikler yaparak çeşitli NLP görevlerinde daha yüksek performans elde etmeyi amaçlamaktadır. GLUE, SQuAD ve RACE gibi standart NLP ölçütlerinde de en iyi sonuçları elde etmiştir (Y. Liu vd., 2019). Bu çalışmada, Türkçe veri kümesiyle eğitilen ve Türkçe doğal dil işleme için uyarlanan RoBERTaTurk modeli kullanılmıştır (byunal, 2023).

3.2.7. Yapay tavşan optimizasyon algoritması

Wang vd., 2022 tarafından sunulan çalışmada, mühendislik optimizasyon problemlerini çözmek amacıyla Yapay Tavşan Optimizasyonu (ARO) adı verilen yeni bir biyo-esinlenmiş meta-sezgisel algoritma önerilmektedir (L. Wang vd., 2022). Algoritma, tavşanların doğadaki hayatta kalma stratejilerinden ilham almaktadır: yırtıcılardan korunmak için yuvalarından uzakta yiyecek arama ("dolambaçlı yiyecek arama" - detour foraging) ve düşmanlardan saklanmak amacıyla rastgele bir yuva seçme

("rastgele saklanma" - random hiding). Bu mekanizmaların modellenmesi sayesinde algoritma, küresel ve yerel arama yeteneklerini önemli ölçüde geliştirerek mühendislik problemlerinin çözümünde başarılı bir performans sergilemiştir. Algoritmanın tez çalışmasında kullanımı Şekil 3.10'da sunulmuştur. Algoritma temel olarak iki doğal davranıştan esinlenmiştir: Dolaylı beslenme ve rastgele saklanma.



Şekil 3.10. ARO algoritması ile oluşturulan kolektif öğrenme modelinin akış şeması (L. Wang vd., 2022'den değiştirilerek alınmıştır)

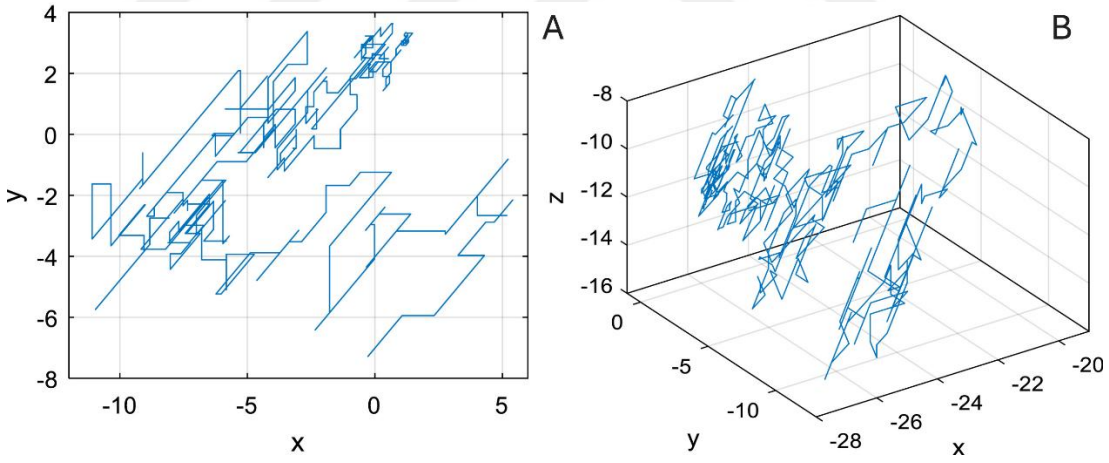
Dolaylı beslenme, tavşanların yırtıcılardan korunmak için yuvalarına yakın otları yemek yerine uzak yerlerdeki otlar ile beslenmesidir. Beslenme hareketi için diğer tavşanların yuvalarının yakınına gider. Bu davranış çözüm uzayında yeni alanların aranmasını temsil eder. Dolaylı beslenme hareketi Denklem 3.1 ile gerçekleştirilir.

$$\vec{v}_i(t+1) = \vec{x}_j(t) + R\left(\vec{x}_i(t) - \vec{x}_j(t)\right) + round\left(0.5\,.\,(0.05 + r_1)\right) \cdot n_1 \qquad (3.1)$$

Dolaylı beslenmede:

- $\vec{x}_i(t)$ i. Tavşanın konumunu
- $\vec{x}_j(t)$ rastgele seçilen başka bir tavşanın konumunu
- R tavşanların koşma operatörünü
- n_1 normal dağılım ile oluşturulan gürültüyü

temsil etmektedir. Denklem 3.1 kullanıldığında tavşanların hareketleri 2 ve 3 boyutlu uzayda Şekil 3.11'deki gibi ortaya çıkmaktadır.



Şekil 3.11. Tavşanların hareketleri **A)** 2-D **B)** 3-D

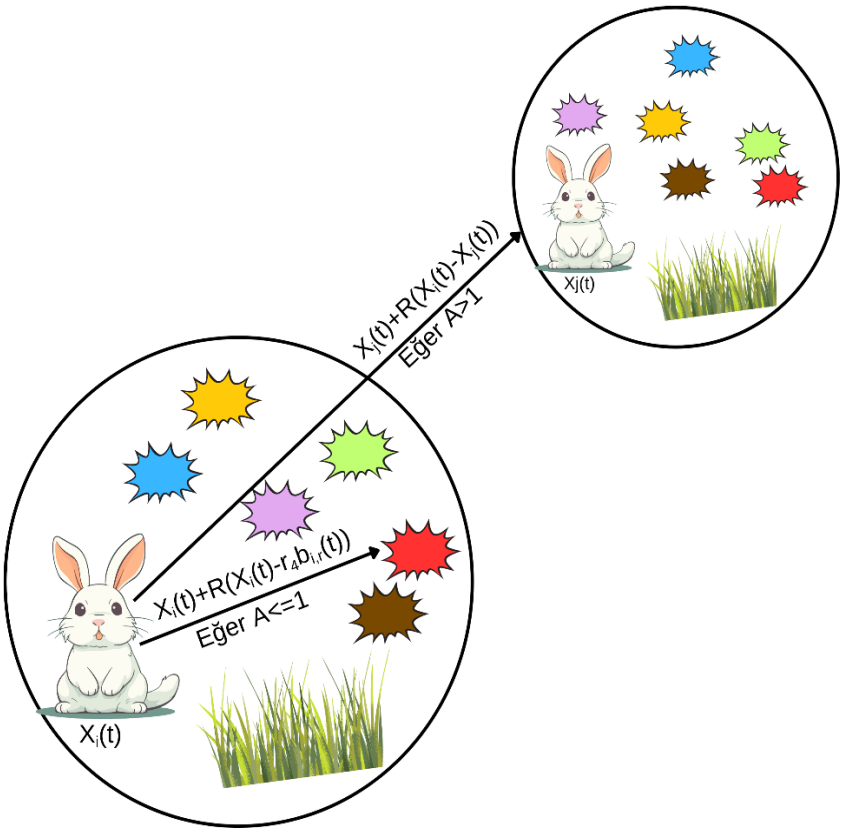
Rastgele saklanma, tavşanların birçok tünel kazması ancak bunlardan birine rastgele gizlenerek avcılardan kaçması davranışıdır. Rastgele saklanma, her tavşanın kendi etrafında d tane gizlenme tüneli kazıp, bunlardan rastgele birini seçerek gizlenmesidir. Bu hareket mevcut iyi çözümlerin yakınında iyileştirme (yerel arama) olarak tanımlanmaktadır. Dolaylı beslenme hareketi Denklem 3.2 ile gerçekleştirilir.

$$\vec{v}_i(t+1) = \vec{x}_i(t) + R\left(r_4, \vec{b}_{i,r}(t) - \vec{x}_i(t)\right) \quad (3.2)$$

Rastgele saklanmada:

- $\vec{v}_i(t)$ tavşan i'nin bir sonraki döngüdeki yeni pozisyonunu
- $\vec{x}_i(t)$ şu anki pozisyonunu
- R tavşanların koşma operatörünü
- r_4 [0,1] aralığında rastgele bir sayıyı
- $b_{i,r}$ tavşan i'nin d adet tünelde rastgele seçilen r. Tüneli

temsil etmektedir. Denklem 3.2 kullanılarak tavşanın kendi tünellerinden birine yönelmesi sağlanarak saklanma işlemi gerçekleştirilir.



Şekil 3.12. Enerji faktörü A'ya dayalı arama mekanizması (L. Wang vd., 2022'den değiştirilerek alınmıştır)

Enerji azaltma stratejisi, tavşanların başlangıçta enerjilerinin yüksek olup daha çok keşif yaparken, zamanla enerjilerinin azalması durumunda saklanma eğilimine başlamayarak, enerji durumuna göre keşif ve sömürü dengesinin sağlanmasıdır. Tavşanlara ait enerji

hesaplaması Denklem 3.3 ile gerçekleştirilir. *A(t)>1* olduğu durumda keşif işlemi yapılırken *A(t)≤1* olduğu durumda ise saklanma işlemi gerçekleştirilir. Keşif ve sömürü dengesi Şekil 3.12'de sunulmuştur.

$$A(t) = 4\left(1 - \frac{t}{T}\right) * \ln\left(\frac{1}{r}\right) \tag{3.3}$$

Yapay tavşan optimizasyon algoritması:

- Tavşanlara ait başlangıç konumunun rastgele belirlenmesi
- Her döngüde, enerji faktörüne göre keşif ve sömürü yapılması
- Daha iyi çözüm varsa yeni pozisyonların kabul edilmesi
- Döngü sayısı tamamlandığında en iyi çözümün geri döndürülmesi

işlem adımlarından oluşmaktadır.

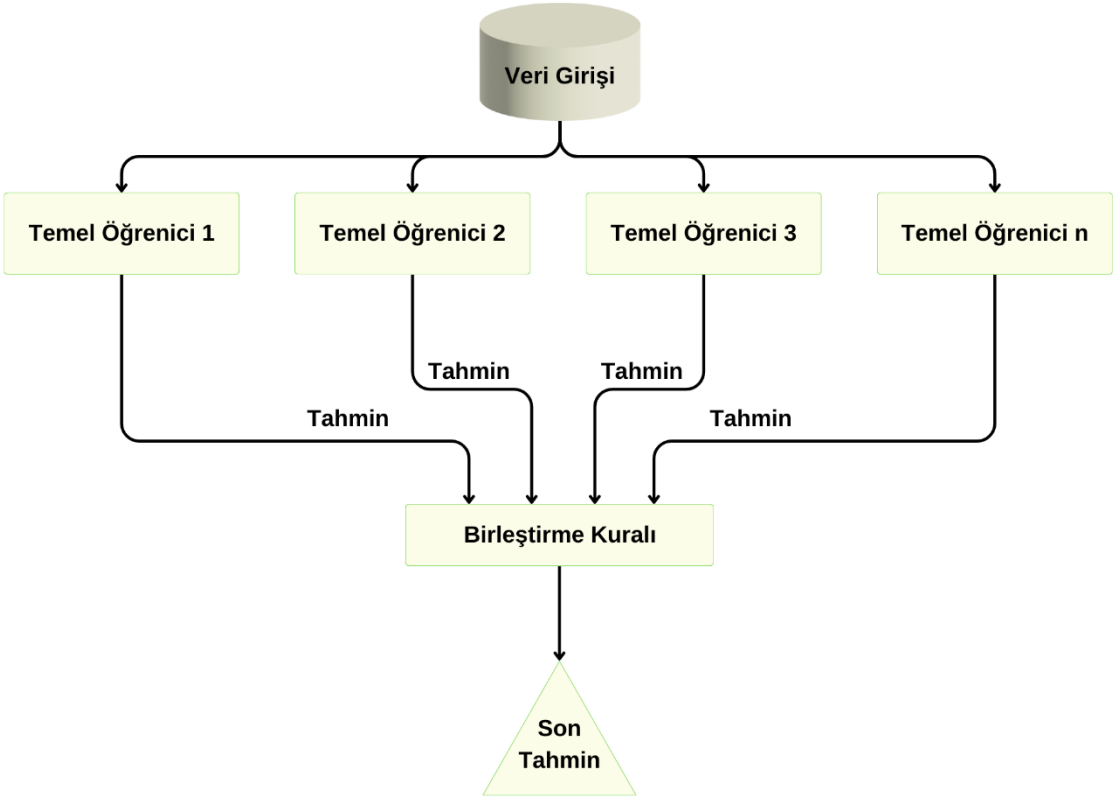
3.2.8. Kolektif öğrenme

Kolektif öğrenme, birden fazla temel modeli birleştirerek bireysel modellerden daha iyi tahmin performansı elde etmeyi hedefleyen bir makine öğrenimi yöntemidir (Mienye ve Swart, 2025). İki veya daha fazla makine öğrenmesi yönteminin bir araya getirerek yalnız kullanıldıklarından daha iyi sonuçlar üretir. Bu yöntem, çok sayıda temel öğreniciyi kullanarak öğrenme sisteminin genelleme yeteneğini artırmayı amaçlar (Yang vd., 2023). Bu yöntem, tahmin performansını ve sağlamlığı artırmak için birden fazla modeli birleştirir (Mienye ve Swart, 2025). Bireysel temel öğrenicilerden daha iyi sonuçlar ve genelleme yeteneği elde etmek için kullanılır (Mienye ve Sun, 2022). Öğrenme sistemlerinin genelleştirilmesini artırmanın önemli avantajları nedeniyle, kolektif öğrenmenin teorik algoritmaları ve uygulamaları üzerine yapılan araştırmalar her zaman araştırma alanında ilgi odağı olmuştur (Yang vd., 2023).

En çok bilinen yöntemler arasında örnekleme ve birleştirme (bagging), artırmalı öğren me (boosting) ve istifleme (stacking) yer alır (Mienye ve Sun, 2022). Örnek olarak, bagging için rastgele orman; boosting için AdaBoost, gradyan artırmalı öğrenme (gradient boosting), aşırı gradyan artırmalı öğrenme (XGBoost), hafif gradyan artırmalı öğrenme (LightGBM) ve kategorik artırmalı öğrenme (CatBoost) gösterilebilir. Büyük

dil modelleri için kullanılan yöntemler arasında model seviyesinde kolektif, parametre seviyesinde kolektif, göreve özgü kolektif, damıtma tabanlı kolektif, uzmanlar karışımı ve hibrit /çoklu ajan kolektif yöntemleri bulunur (Mienye ve Swart, 2025). Geleneksel kolektif öğrenme için ise heterojen ve homojen temel öğrenici üretimi, ağırlıklandırma yöntemleri, meta-öğrenme yöntemleri ve kolektif seçim yöntemleri öne çıkmaktadır (Yang vd., 2023).

LightGBM, büyük verileri hızlı işler ve yüksek doğruluk sağlar; gradyan temelli tek yönlü örnekleme (GOSS) ve birbirini dışlayan özelliklerin gruplanması (EFB) teknikleriyle diğer boosting yöntemlerinden daha iyi doğruluk sunar, ancak küçük veri kümelerinde aşırı uyum riski taşır (Mienye ve Sun, 2022). Büyük dil modelleri kullanılarak oluşturulan kolektif öğrenme yöntemlerinden biri olan model seviyesinde yöntemler yüksek tahmin performansı ve sağlam genelleme sunar. Bu yöntemde, yüksek hesaplama maliyeti, gecikme ve düşük yorumlanabilirlik gibi olumsuzluklar dezavantajları olarak nitelendirilmektedir. Parametre seviyesinde yöntemler hesaplama verimliliği ve ölçeklenebilirlik sağlasa da model çeşitliliği sınırlı kalır ve yorumlanabilirlik zorlaşır. Göreve özgü kolektif yöntemler, yüksek alan uzmanlığı, ölçeklenebilirlik ve hedeflenen uygulamalara uyarlanabilirlik gibi avantajlar sunarken, temel sınırlamaları girdi kalitesine bağımlı olmalarıdır. Çok modlu entegrasyon yeteneği ve çapraz modlu akıl yürütme ile heterojen veri ortamlarında sağlamlık gibi özellikler, çok modlu LLM kolektif öğrenme yöntemleri için belirgin avantajlardır (Mienye ve Swart, 2025). Geleneksel kolektif öğrenme (TEL) yöntemlerinde, birden fazla temel öğreniciyi eğitip birleştirdiği için tek bir temel öğreniciye kıyasla zaman ve alan maliyetleri artmaktadır. Bu durum, kolektif öğrenmenin yaygınlaşmasını engelleyen bir faktördür. Ancak, modern derin sinir ağlarının çok büyük sayıda parametreye sahip olması nedeniyle, derin öğrenme topluluklarının (EDL) zaman ve alan maliyetleri geleneksel kolektif öğrenmeden çok daha fazladır. Bu yüksek maliyet, EDL'nin gelişiminin TEL'e göre geride kalmasının ana nedenlerinden biridir (Yang vd., 2023). Kolektif öğrenmenin temel öğrenici modeller tarafından gerçekleştirilmesi Şekil 3.13'te sunulmuştur.



Şekil 3.13. Kolektif öğrenme gösterimi

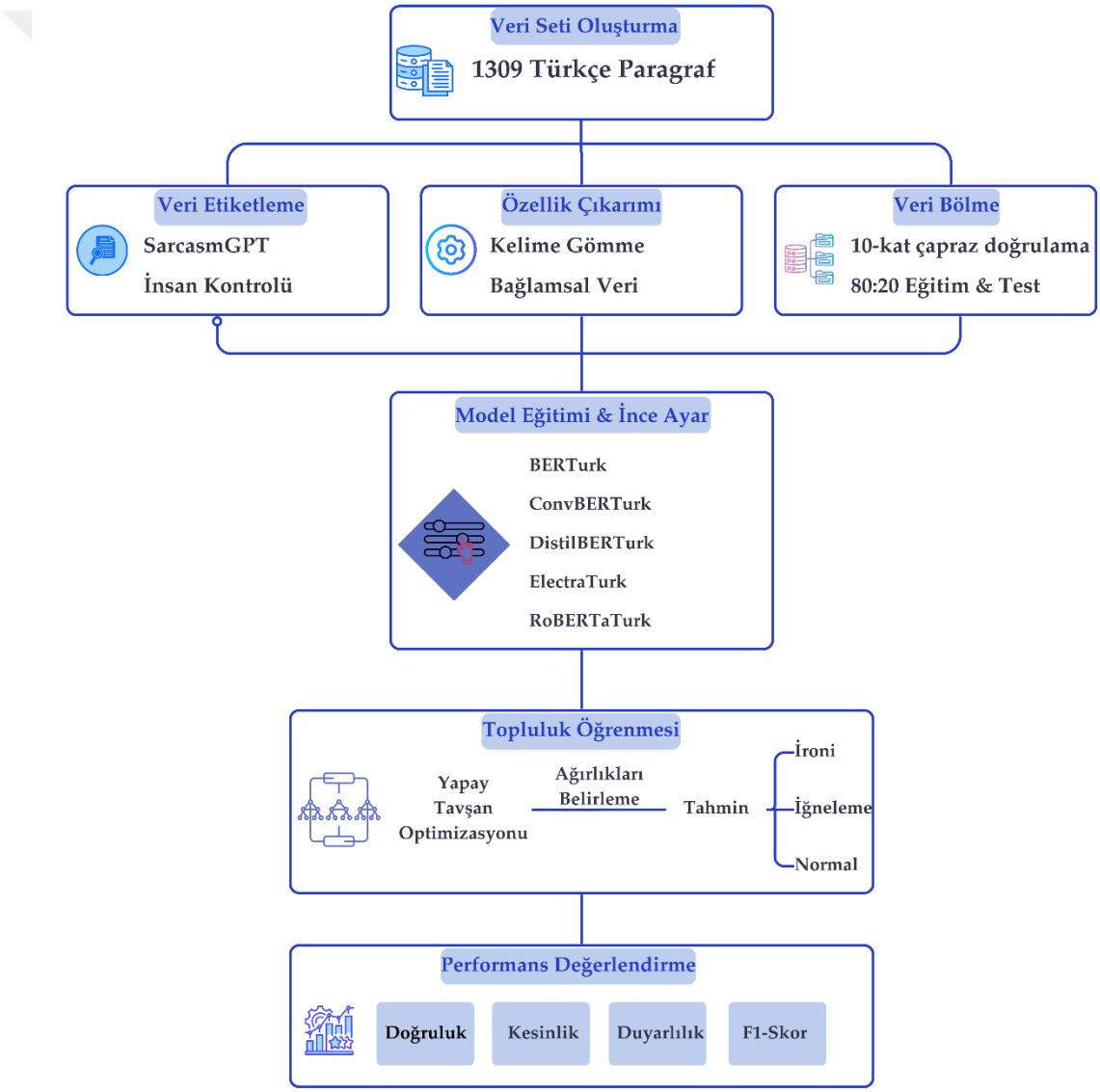
Büyük dil modellerinde (LLM) kolektif öğrenme, birden fazla dil modelinin birleştirilmesiyle tahminsel performansı ve sağlamlığı artırmak için kullanılır ve model seviyesinde kolektif öğrenme farklı dil modellerinin çıktılarını birleştirerek yüksek tahmin performansı, genelleştirme yeteneğinin artması, varyans azalması avantajlarını sağlar. Bu yöntemler, özellikle tıbbi soru-cevap uygulamalarında doğruluk artışı ve yazılım hata lokalizasyonunda daha güvenilir hata tespiti sunar. Ayrıca, ön yargıları azaltma ve yorumlanabilirlik gibi zorlukları aşmada önemlidir. Model seviyesinde kolektif öğrenme yöntemleri, farklı dil çiftleri veya görevler için uzmanlaşmış modelleri birleştirerek makine çevirisinde olduğu gibi daha tutarlı ve doğru çeviriler üretir. Çok modlu kolektif öğrenme yöntemleri, metin ve görüntü gibi farklı veri türlerini birlikte işleyerek çapraz akıl yürütme yeteneğini güçlendirir ve heterojen veri ortamlarında sağlamlık sağlar. Ancak, hizalama, senkronizasyon ve ölçeklenebilirlik gibi zorluklarla karşılaşır. Parametre seviyesindeki kolektif öğrenme yöntemleri, aynı LLM mimarisinin farklı alternatiflerini kullanarak hesaplama verimliliği, ortalama doğruluk artışı ve ölçeklenebilirlik sağlar. Göreve özgü kolektif öğrenme yöntemleri, özel uygulamaların

gereksinimlerini karşılamak üzere belirli alt görevler veya alanlar için ince ayar yapılmış LLM'leri entegre ederek yüksek alan uzmanlığı, ölçeklenebilirlik ve hedeflenen uygulamalara uyarlanabilirlik gibi avantajlar sunar. Modüler yapıları, disiplinler arası alanlarda alan bilgisinin verimli entegrasyonuna olanak tanır. Damıtma temelli kolektif öğrenme, büyük ve karmaşık öğretici modellerin bilgisini daha küçük, verimli öğrenci modellere aktararak hem yüksek hesaplama verimliliği hem de kaynakların sınırlı olduğu ortamlarda düşük gecikmeyle çalışmayı sağlar. Uzmanlar karışımı (Mixture-of-Experts - MoE) yaklaşımı, girdileri uzmanlaşmış modellere dinamik olarak yönlendirerek ölçeklenebilirlik ve verimli kaynak kullanımı gibi avantajları bulunmaktadır. Karmaşık yönlendirme mekanizmaları ve modellere yüklerin dengelemesi gibi dezavantajları bulunur. Hibrit ajan kolektif öğrenme, birden fazla farklı LLM ajanının iş birliği yaparak karmaşık, çok yönlü görevleri ele almasını sağlar ve gelişmiş karar alma yetenekleri ile dinamik ortamlarda uyarlanabilirlik sunar. Ancak koordinasyon karmaşıklığının fazla olması, modellere gönderilen görevlerin senkronizasyon yükü gibi zorluklar getirir. Büyük dil modellerinde kolektif öğrenme yaklaşımları, sağlık ve hukuk gibi riskli alanlarda güvenilirlik ve şeffaflık gereksinimlerini karşılamak için avantajlıdır. Bunu yanı sıra genel performansın iyileştirmesi açısından da faydalıdır. Yüksek hesaplama maliyeti, yüksek gecikme süresi ve düşük yorumlanabilirlik gibi genel sınırlamalar devam etmektedir (Mienye ve Swart, 2025).

3.3. Önerilen Yöntem

Bu çalışmada, Türkçe doğal dil işleme görevleri için BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk ve RoBERTaTurk modelleri, özel olarak oluşturulan bir veri kümesi kullanılarak yeniden eğitilmiştir. Her bir model, farklı veri kümelerinde çalıştırılarak eğitim işlemi tamamlanmış ve test veri kümesi üzerinde sonuçları alınmıştır. Ardından, yapay tavşan optimizasyon (ARO) algoritması kullanılarak modellerin tahmin başarısına göre ağırlıkları belirlenmiştir. Bu algoritma, biyolojik olarak esinlenmiş bir meta-sezgisel yaklaşım benimseyerek, modellerin kolektif öğrenmedeki kombinasyonunu optimize etmiş ve her modelin katkısını ağırlıklandırmıştır. Son aşamada, bu ağırlıklandırılmış kolektif öğrenme modeli ile yeni tahminler gerçekleştirilmiştir. Önerilen kolektif öğrenme Şekil 3.14'te sunulmuştur.

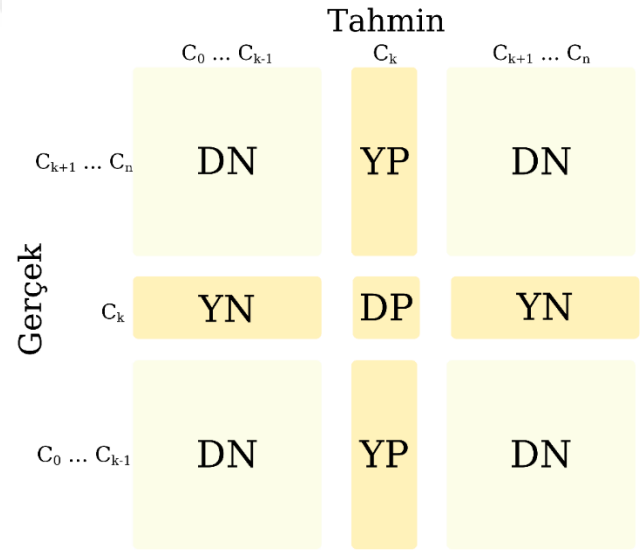
Kolektif öğrenmenin bireysel modellere göre performans artışı sağlaması nedeniyle bireysel modeller ayrıca kolektif öğrenmeye dahil edilerek performans artışı amaçlanmaktadır. Oluşturulan veri kümeleri ile ön eğitimli modeller eğitilerek ironi ve iğneleme için modeller oluşturulur. Ardından yapay tavşan algoritmasında parametreler belirlenerek başlangıç aşamasında çalıştırılır. Belirlenen döngü sayısı boyunca modellerin doğru tahmin performansları doğrultusunda ağırlıkları güncellenir. Döngü sayısı bittiğinde her model için nihai bir ağırlık belirlenir. Kolektif model belirlenen ağırlıklara göre yeni tahminde bulunur. Son aşamada modellerin bireysel olarak yaptıkları tahminler ve kolektif öğrenme tahmini analiz edilir.



Şekil 3.14. Önerilen kolektif öğrenme yöntemi

3.4. İroni ve İğneleme Tespitinde Kullanılan Performans Metrikleri

Makine öğrenmesi problemlerinde kullanılan yöntemlerin değerlendirilmesinde, yöntemin performans açısından analizine ihtiyaç duyulmaktadır. Analiz sonucunda kullanılan yöntemin diğer yöntemlerle karşılaştırması yapılarak etkinliği konusunda sonuca varılır. Bu çalışmada Türkçe metinlerde ön eğitimli dil modellerinin ve bu modellerden elde etilen kolektif öğrenmenin analizinde literatürde de sık kullanılan doğruluk, kesinlik, duyarlılık ve F1-skor metrikleri kullanılmıştır. Kullanılan metriklerin hesaplanmasında karmaşıklık matrisi kullanılmış olup Şekil 3.15'te sunulmuştur.



Şekil 3.15. Çok sınıflı karmaşıklık matrisi

Karmaşıklık matrisi kullanılarak, gerçekte pozitif etiketine sahip olan örneğin, pozitif olarak tahmin edilmesi Doğru Pozitif (DP) olarak ifade edilmektedir. İroni sınıfı ele alındığında, gerçekte ironik olan bir ifadenin model tarafından da ironik olarak tahmin edilmesi anlamına gelir. Gerçekte negatif etiketine sahip olan örneğin, pozitif olarak tahmin edilmesi Yanlış Pozitif (YP) olarak ifade edilmektedir. Bu gerçekte ironik olmayan bir ifadenin model tarafından yanlış şekilde ironik olarak etiketlenmesi durumudur. Gerçekte pozitif etiketine sahip olan örneğin, negatif olarak tahmin edilmesi Yanlış Negatif (YN) olarak ifade edilmektedir. Bu gerçekte ironik olan bir ifadenin model tarafından ironik değil şeklinde değerlendirilmesidir. Gerçekte negatif etiketine

sahip olan örneğin, negatif olarak tahmin edilmesi Doğru Negatif (DN) olarak ifade edilmektedir. Bu gerçekte ironik olmayan bir ifadenin model tarafından ironik değil şeklinde değerlendirilmesidir.

Karmaşıklık matrisi ve örneklerin sınıflandırılması sonucu elde edilen DP, YP, YN ve DN değerlerinin kullanılmasıyla performans ölçüm metrikleri olan doğruluk, kesinlik, duyarlılık ve F1-skor değerleri elde edilmiştir.

Doğruluk, doğru tahmin edilen pozitif örneklerin, gerçekte pozitif olan örneklere oranını vermektedir. Doğruluk ölçüm metriğinin hesaplanması için Denklem 3.4 kullanılmıştır.

$$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN} \tag{3.4}$$

Kesinlik, pozitif olarak tahmin edilen örneklerin, gerçekte ne kadarının pozitif olduğunu belirtir. Kesinlik ölçüm metriğinin hesaplanması için Denklem 3.5 kullanılmıştır.

$$Kesinlik = \frac{DP}{DP + YP} \tag{3.5}$$

Duyarlılık, pozitif olarak tahmin edilen örneklerin, gerçekte pozitif olan örneklere oranını vermektedir. Duyarlılık ölçüm metriğinin hesaplanması için Denklem 3.6 kullanılmıştır.

$$Duyarlılık = \frac{DP}{DP + YN} \tag{3.6}$$

F1-skor, kesinlik ve duyarlılık metriklerinin harmonik ortalamasıdır. F1-skor ölçüm metriğinin hesaplanması için Denklem 3.7 kullanılmıştır.

$$F1 - skor = \frac{2 \cdot (Kesinlik \cdot Duyarlılık)}{Kesinlik + Duyarlılık} \tag{3.7}$$

4. BULGULAR

Çalışmanın bu bölümünde, köşe yazıları incelenerek oluşturulan bağlam temelli 5 adet veri kümesi üzerinde gerçekleştirilen önceden eğitilmiş dil modelleri ve klasik makine öğrenmesi algoritmasına ait deneysel çalışmaların sonuçları analiz edilmiştir. Model olarak BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk, RoBERTaTurk ve SVM yöntemleri kullanılmıştır. Ayrıca, bu modellerin çıktılarını birleştirmek için bir yapay tavşan optimizasyon yöntemi kullanılmış ve bireysel modellerin güçlü yönlerinden yararlanan kolektif öğrenme ile nihai bir tahmin üretilmiştir. Çalışma süresince bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümesi mimarileri üzerinde çeşitli analizler gerçekleştirilmiştir. Modellerin sınıflandırma performansı, verilen görevdeki etkinliklerinin ayrıntılı bir değerlendirmesini sağlayan doğruluk, kesinlik, duyarlılık ve F1-skor performans metrikleri kullanılarak kapsamlı bir şekilde değerlendirilmiştir.

Gerçekleştirilen deneysel çalışmalar 2 aşamalı olarak ele alınmıştır. İlk aşamada modeller bireysel olarak analiz edilmiş ve 10 kat çapraz doğrulama yöntemi kullanılmıştır. İkinci aşamada ön eğitimli dil modelleri kolektif öğrenmenin bir parçası olarak kullanılmış ve 80:20 oranında eğitim-test bölünmesi kullanılmıştır. Her iki aşamada yapılan deneyler https://www.kaggle.com adresi üzerinde GPU P100 ve GPU T4 grafik işlemcili sistem üzerinde gerçekleştirilmiştir. Dönüştürücü modellerinde ince ayar sırasında maksimum uzunluk 512, küçük harf, kesme ve ekleme parametreleri evet, öğrenme katsayısı 0,00002-0,00005 aralığında, kayıp yöntemi çapraz entropi kaybı, optimizasyon yöntemi AdamW olarak kullanılmıştır. SVM yönteminde çekirdek fonksiyon için rbf ve linear, düzenlileştirme için 0,01-100 arasında onar kat arayla, çekirdek katsayısı için scale, auto, 0,1-0,001 aralığında ızgara yöntemi ile uygun parametreler aranmıştır.

4.1. Modellerin Bireysel Performans Değerlendirmesi

Çalışmanın bu kısmında BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk, RoBERTaTurk ve SVM yöntemlerinin bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümeleri üzerinde bireysel sınıflandırma performansına ait sonuçlar

değerlendirilmiştir. Modellerin bireysel değerlendirmesinde 10 kat çapraz doğrulama kullanılıp sonuçlar doğruluk, kesinlik, duyarlılık ve F1-skor performans metrikleri ile değerlendirilmiştir. Modellerin veri kümelerine göre performans sonuçları Çizelge 4.1-4.13 arasında sunulmuştur.

Çizelge 4.1'de bir-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda en başarılı model %65,63 doğruluk, %63,19 kesinlik, %61,89 duyarlılık ve %62,58 F1-skor metrik değerleri ile BERTurk olurken, SVM modelinin GloVe gömme yöntemi ile kullanılması durumunda başarı açısından %40,50 doğruluk, %38,50 kesinlik, %40,50 duyarlılık ve %39,47 F1-skor metrik değerleri ile son sırada yer almaktadır.

Çizelge 4.1. Modellerin bir-gram veri kümesinde bireysel performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	bir-gram	**65,63**	**63,19**	**61,89**	**62,58**
ConvBERTurk	bir-gram	60,89	59,56	58,19	58,88
DistilBERTurk	bir-gram	57,45	55,61	54,89	55,22
ELECTRATurk	bir-gram	59,59	55,30	56,56	55,87
RoBERTaTurk	bir-gram	54,32	50,93	50,72	50,80
SVM + TF-IDF	bir-gram	45,80	44,20	45,80	44,98
SVM + GloVe	bir-gram	40,50	38,50	40,50	39,47

Modellerin bir-gram veri kümesinde sınıflara göre elde ettiği performans metrikleri Çizelge 4.2'de sunulmuştur. Sınıflara göre modellerin performansları değerlendirildiğinde, ironi sınıfı için BERTurk modeli %69 doğruluk ile en iyi model olurken, SVM ile GloVe kelime gömme yöntemlerinin birlikte kullanıldığı yöntem %60 doğruluk ile son sırada yer almaktadır. İğneleme sınıfı için BERTurk modeli %72 doğruluk ile en iyi, SVM ile GloVe yöntemi %66 doğruluk performansı ile son sırada yer almaktadır. Normal sınıfı için BERTurk modeli %89 doğruluk ile en iyi, SVM ile GloVe yöntemi %54 doğruluk performansı ile son sırada yer almaktadır. bir-gram veri kümesinde sınıflara göre performanslara bakıldığında BERTurk modeli en başarılı

model olurken, her üç sınıfta da SVM ile GloVe yöntemi diğer modellere göre başarı oranı daha düşüktür.

Çizelge 4.2. Modellerin bir-gram veri kümesinde sınıflara göre performans sonuçları

Model	Sınıf	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	İroni	**69,00**	52,00	66,00	58,00
	İğneleme	**72,00**	51,00	30,00	38,00
	Normal	**89,00**	85,00	88,00	86,00
ConvBERTurk	İroni	66,00	50,00	61,00	55,00
	İğneleme	71,00	48,00	34,00	40,00
	Normal	83,00	80,00	79,00	79,00
DistilBERTurk	İroni	65,00	47,00	45,00	46,00
	İğneleme	69,00	45,00	41,00	43,00
	Normal	80,00	74,00	78,00	76,00
ELECTRATurk	İroni	66,00	49,00	59,00	54,00
	İğneleme	69,00	34,00	31,00	32,00
	Normal	83,00	81,00	78,00	79,00
RoBERTaTurk	İroni	66,00	47,00	48,00	47,00
	İğneleme	69,00	40,00	25,00	31,00
	Normal	72,00	64,00	77,00	70,00
SVM + TF-IDF	İroni	61,00	40,00	39,00	39,00
	İğneleme	67,00	35,00	22,00	27,00
	Normal	62,00	53,00	67,00	59,00
SVM + GloVe	İroni	60,00	35,00	26,00	30,00
	İğneleme	66,00	32,00	22,00	26,00
	Normal	54,00	45,00	64,00	53,00

Çizelge 4.3'te iki-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda en başarılı model %66,72 doğruluk, %64,72 kesinlik, %63,50 duyarlılık ve %64,10 F1-skor metrik değerleri ile BERTurk olurken, SVM modelinin GloVe gömme yöntemi ile kullanılması durumunda başarı açısından %41,40 doğruluk, %41,60 kesinlik, %41,40 duyarlılık ve %41,50 F1-skor metrik değerleri ile son sırada yer almaktadır.

Çizelge 4.3. Modellerin iki-gram veri kümesinde bireysel performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	iki-gram	**66,72**	**64,72**	**63,50**	**64,10**
ConvBERTurk	iki-gram	65,30	64,39	62,00	63,17
DistilBERTurk	iki-gram	60,91	57,81	57,52	57,66
ELECTRATurk	iki-gram	65,53	64,08	61,93	62,99
RoBERTaTurk	iki-gram	61,24	60,99	56,83	58,84
SVM + TF-IDF	iki-gram	49,10	46,80	49,10	47,93
SVM + GloVe	iki-gram	41,40	41,60	41,40	41,50

Modellerin iki-gram veri kümesinde sınıflara göre elde ettiği performans metrikleri Çizelge 4.4'te sunulmuştur. Sınıflara göre modellerin performansları değerlendirildiğinde, ironi sınıfı için BERTurk modeli %71 doğruluk ile en iyi model olurken, SVM ile GloVe kelime gömme yöntemlerinin birlikte kullanıldığı yöntem %59 doğruluk ile son sırada yer almaktadır. İğneleme sınıfı için BERTurk modeli %75 doğruluk ile en iyi, SVM ile GloVe yöntemi %64 doğruluk performansı ile son sırada yer almaktadır. Normal sınıfı için ELECTRATurk modeli %88 doğruluk ile en iyi, SVM ile GloVe yöntemi %58 doğruluk performansı ile son sırada yer almaktadır. iki-gram veri kümesinde sınıflara göre performanslara bakıldığında BERTurk modeli genellikle en başarılı model olurken, her üç sınıfta da SVM ile GloVe yönteminin diğer modellere göre başarı oranı daha düşüktür.

Çizelge 4.4. Modellerin iki-gram veri kümesinde sınıflara göre performans sonuçları

Model	Sınıf	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	İroni	**71,00**	54,00	62,00	58,00
	İğneleme	**75,00**	55,00	43,00	48,00
	Normal	86,00	83,00	84,00	84,00
ConvBERTurk	İroni	69,00	52,00	71,00	60,00
	İğneleme	74,00	54,00	33,00	41,00
	Normal	86,00	86,00	80,00	83,00
DistilBERTurk	İroni	65,00	47,00	58,00	52,00
	İğneleme	71,00	44,00	34,00	38,00
	Normal	84,00	81,00	80,00	81,00
ELECTRATurk	İroni	65,00	47,00	58,00	52,00
	İğneleme	73,00	52,00	34,00	41,00
	Normal	**88,00**	89,00	83,00	86,00
RoBERTaTurk	İroni	68,00	50,00	67,00	57,00
	İğneleme	74,00	59,00	17,00	26,00
	Normal	81,00	75,00	81,00	78,00
SVM + TF-IDF	İroni	60,00	39,00	42,00	40,00
	İğneleme	67,00	31,00	22,00	26,00
	Normal	70,00	62,00	71,00	66,00
SVM + GloVe	İroni	59,00	38,00	39,00	38,00
	İğneleme	64,00	32,00	31,00	31,00
	Normal	58,00	50,00	49,00	50,00

Çizelge 4.5'te üç-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda en başarılı model %67,42 doğruluk, %64,73 kesinlik, %64,32 duyarlılık ve %64,52 F1-skor metrik değerleri ile BERTurk olurken, SVM modelinin GloVe gömme yöntemi ile kullanılması durumunda başarı açısından %45,10 doğruluk, %43,90 kesinlik, %45,10 duyarlılık ve %44,49 F1-skor metrik değerleri ile son sırada yer almaktadır.

Çizelge 4.5. Modellerin üç-gram veri kümesinde bireysel performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	üç-gram	**67,42**	**64,73**	**64,32**	**64,52**
ConvBERTurk	üç-gram	65,62	61,58	61,97	61,77
DistilBERTurk	üç-gram	62,71	60,23	59,63	59,93
ELECTRATurk	üç-gram	66,17	63,74	62,96	63,35
RoBERTaTurk	üç-gram	61,16	56,84	56,61	56,72
SVM + TF-IDF	üç-gram	52,10	51,20	52,10	51,65
SVM + GloVe	üç-gram	45,10	43,90	45,10	44,49

Modellerin üç-gram veri kümesinde sınıflara göre elde ettiği performans metrikleri Çizelge 4.6'da sunulmuştur. Sınıflara göre modellerin performansları değerlendirildiğinde, ironi sınıfı için BERTurk modeli %70 doğruluk ile en iyi model olurken, SVM ile GloVe kelime gömme yöntemlerinin birlikte kullanıldığı yöntem %60 doğruluk ile son sırada yer almaktadır. İğneleme sınıfı için BERTurk modeli %73 doğruluk ile en iyi, SVM ile GloVe yöntemi %66 doğruluk performansı ile son sırada yer almaktadır. Normal sınıfı için ConvBERTurk modeli %92 doğruluk ile en iyi, SVM ile GloVe yöntemi %62 doğruluk performansı ile son sırada yer almaktadır. üç-gram veri kümesinde sınıflara göre performanslara bakıldığında BERTurk modeli genellikle en başarılı model olurken, her üç sınıfta da SVM ile GloVe yönteminin diğer modellere göre başarı oranı daha düşüktür.

Çizelge 4.6. Modellerin üç-gram veri kümesinde sınıflara göre performans sonuçları

Model	Sınıf	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	İroni	**70,00**	53,00	58,00	55,00
	İğneleme	**73,00**	51,00	45,00	48,00
	Normal	90,00	89,00	89,00	89,00
ConvBERTurk	İroni	69,00	52,00	71,00	60,00
	İğneleme	70,00	44,00	47,00	45,00
	Normal	**92,00**	89,00	92,00	90,00
DistilBERTurk	İroni	68,00	50,00	46,00	48,00
	İğneleme	71,00	49,00	42,00	45,00
	Normal	88,00	84,00	87,00	85,00
ELECTRATurk	İroni	65,00	46,00	48,00	47,00
	İğneleme	72,00	49,00	47,00	48,00
	Normal	90,00	90,00	87,00	89,00
RoBERTaTurk	İroni	69,00	51,00	53,00	52,00
	İğneleme	73,00	43,00	17,00	24,00
	Normal	84,00	79,00	83,00	81,00
SVM + TF-IDF	İroni	61,00	41,00	43,00	42,00
	İğneleme	68,00	38,00	27,00	32,00
	Normal	73,00	67,00	74,00	70,00
SVM + GloVe	İroni	60,00	39,00	39,00	39,00
	İğneleme	66,00	33,00	23,00	27,00
	Normal	62,00	54,00	62,00	58,00

Çizelge 4.7'de dört-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda en başarılı model %68,76 doğruluk, %66,29 kesinlik, %65,19 duyarlılık ve %65,75 F1-skor metrik değerleri ile ConvBERTurk olurken, SVM modelinin GloVe gömme yöntemi ile kullanılması durumunda başarı açısından %45,30 doğruluk, %44,30 kesinlik, %45,30 duyarlılık ve %44,80 F1-skor metrik değerleri ile son sırada yer almaktadır.

Çizelge 4.7. Modellerin dört-gram veri kümesinde bireysel performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	dört-gram	67,07	65,08	63,99	64,50
ConvBERTurk	dört-gram	**68,76**	**66,29**	**65,19**	**65,75**
DistilBERTurk	dört-gram	66,39	62,70	62,40	62,57
ELECTRATurk	dört-gram	67,00	62,22	63,72	62,93
RoBERTaTurk	dört-gram	61,11	57,62	56,51	57,05
SVM + TF-IDF	dört-gram	53,90	53,10	53,90	53,50
SVM + GloVe	dört-gram	45,30	44,30	45,30	44,80

Modellerin dört-gram veri kümesinde sınıflara göre elde ettiği performans metrikleri Çizelge 4.8'de sunulmuştur. Sınıflara göre modellerin performansları değerlendirildiğinde, ironi sınıfı için ConvBERTurk modeli %70 doğruluk ile en iyi model olurken, SVM ile GloVe kelime gömme yöntemlerinin birlikte kullanıldığı yöntem %60 doğruluk ile son sırada yer almaktadır. İğneleme sınıfı için DistilBERTurk modeli %73 doğruluk ile en iyi, SVM ile GloVe yöntemi %64 doğruluk performansı ile son sırada yer almaktadır. Normal sınıfı için ConvBERTurk modeli %94 doğruluk ile en iyi, SVM ile GloVe yöntemi %65 doğruluk performansı ile son sırada yer almaktadır. dört-gram veri kümesinde sınıflara göre performanslara bakıldığında ConvBERTurk modeli genellikle en başarılı model olurken, her üç sınıfta da SVM ile GloVe yönteminin diğer modellere göre başarı oranı daha düşüktür.

Çizelge 4.8. Modellerin dört-gram veri kümesinde sınıflara göre performans sonuçları

Model	Sınıf	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	İroni	69,00	54,00	61,00	57,00
	İğneleme	71,00	48,00	42,00	45,00
	Normal	92,00	92,00	88,00	90,00
ConvBERTurk	İroni	**70,00**	54,00	63,00	57,00
	İğneleme	72,00	52,00	38,00	41,00
	Normal	**94,00**	92,00	93,00	96,00
DistilBERTurk	İroni	69,00	53,00	60,00	57,00
	İğneleme	**73,00**	50,00	33,00	40,00
	Normal	89,00	83,00	92,00	87,00
ELECTRATurk	İroni	69,00	54,00	56,00	55,00
	İğneleme	71,00	41,00	43,00	42,00
	Normal	92,00	90,00	91,00	91,00
RoBERTaTurk	İroni	65,00	47,00	68,00	56,00
	İğneleme	72,00	46,00	15,00	22,00
	Normal	85,00	78,00	86,00	82,00
SVM + TF-IDF	İroni	62,00	44,00	48,00	45,00
	İğneleme	68,00	37,00	27,00	30,00
	Normal	77,00	70,00	76,00	72,00
SVM + GloVe	İroni	60,00	40,00	39,00	48,00
	İğneleme	64,00	31,00	27,00	29,00
	Normal	65,00	56,00	62,00	59,00

Çizelge 4.9'da tüm-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda en başarılı model %70,74 doğruluk, %67,74 duyarlılık metriklerinde ConvBERTurk, %68,35 kesinlik ve %67,90 F1-skor metriklerinde BERTurk modeli olmuştur. SVM modelinin GloVe gömme yöntemi ile kullanılması durumunda başarı açısından %46,10 doğruluk, %45,90 kesinlik, %46,10 duyarlılık ve %46 F1-skor metrik değerleri ile son sırada yer almaktadır.

Çizelge 4.9. Modellerin tüm-gram veri kümesinde bireysel performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	tüm-gram	70,28	**68,35**	67,40	**67,90**
ConvBERTurk	tüm-gram	**70,74**	67,99	**67,74**	67,84
DistilBERTurk	tüm-gram	66,47	63,22	63,20	63,22
ELECTRATurk	tüm-gram	67,76	65,05	64,87	64,96
RoBERTaTurk	tüm-gram	61,88	52,00	57,03	54,39
SVM + TF-IDF	tüm-gram	58,30	57,80	58,30	58,05
SVM + GloVe	tüm-gram	46,10	45,90	46,10	46,00

Modellerin tüm-gram veri kümesinde sınıflara göre elde ettiği performans metrikleri Çizelge 4.10'da sunulmuştur. Sınıflara göre modellerin performansları değerlendirildiğinde, ironi sınıfı için ConvBERTurk ve BERTurk %72 doğruluk ile en iyi modeller olurken, SVM ile GloVe kelime gömme yöntemlerinin birlikte kullanıldığı yöntem %60 doğruluk ile son sırada yer almaktadır. İğneleme sınıfı için BERTurk modeli %74 doğruluk ile en iyi, SVM ile GloVe yöntemi %64 doğruluk performansı ile son sırada yer almaktadır. Normal sınıfı için ConvBERTurk modeli %97 doğruluk ile en iyi, SVM ile GloVe yöntemi %68 doğruluk performansı ile son sırada yer almaktadır. tüm-gram veri kümesinde sınıflara göre performanslara bakıldığında ConvBERTurk modeli genellikle en başarılı model olurken, her üç sınıfta da SVM ile GloVe yönteminin diğer modellere göre başarı oranı daha düşüktür.

Çizelge 4.10. Modellerin tüm-gram veri kümesinde sınıflara göre performans sonuçları

Model	Sınıf	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	İroni	**72,00**	57,00	63,00	60,00
	İğneleme	**74,00**	54,00	47,00	51,00
	Normal	94,00	94,00	91,00	93,00
ConvBERTurk	İroni	**72,00**	58,00	53,00	56,00
	İğneleme	73,00	51,00	53,00	52,00
	Normal	**97,00**	95,00	97,00	96,00
DistilBERTurk	İroni	68,00	52,00	54,00	53,00
	İğneleme	72,00	47,00	42,00	45,00
	Normal	93,00	90,00	94,00	92,00
ELECTRATurk	İroni	69,00	54,00	50,00	52,00
	İğneleme	71,00	49,00	49,00	49,00
	Normal	95,00	92,00	95,00	94,00
RoBERTaTurk	İroni	68,00	46,00	64,00	54,00
	İğneleme	72,00	31,00	16,00	21,00
	Normal	84,00	79,00	91,00	85,00
SVM + TF-IDF	İroni	64,00	47,00	55,00	50,00
	İğneleme	69,00	41,00	32,00	36,00
	Normal	83,00	79,00	79,00	79,00
SVM + GloVe	İroni	60,00	40,00	42,00	41,00
	İğneleme	64,00	32,00	31,00	31,00
	Normal	68,00	60,00	60,00	60,00

Çizelge 4.1-4.10'da bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümeleri üzerinde BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk, RoBERTaTurk ve SVM (TF-IDF ve GloVe gömme yöntemleri) modellerinin 10 kat çapraz doğrulama sonuçları verilerek her bir modelin farklı veri kümeleri üzerindeki performans analizleri verilmiştir. Yapılan analizlerin değerlendirilmesi için veri kümelerine göre en iyi performansı sunan modellerin sonuçları Çizelge 4.11'de sunulmuştur.

Çizelge 4.11. Veri kümelerine göre en iyi performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	bir-gram	65,63	63,19	61,89	62,58
BERTurk	iki-gram	66,72	64,72	63,50	64,10
BERTurk	üç-gram	67,42	64,73	64,32	64,52
ConvBERTurk	dört-gram	68,76	66,29	65,19	65,75
ConvBERTurk	tüm-gram	**70,74**	**67,99**	**67,74**	**67,84**

Performans metriklerine göre BERTurk, bir-gram veri kümesinde %65,63 doğruluk değeri, iki-gram veri kümesinde %66,72 doğruluk değeri, üç-gram veri kümesinde %67,42 doğruluk değeri ile en başarılı model olmuştur. ConvBERTurk, dört-gram veri kümesinde %68,76 doğruluk değeri ve tüm-gram veri kümesinde %70,74 doğruluk değeri ile en başarılı model olmuştur. BERTurk modelinin bir-gram veri kümesinde elde ettiği doğruluk, kesinlik, duyarlılık ve F1-Skor metrik değerleri ile ConvBERTurk modelinin tüm-gram veri kümesinde elde ettiği metrik değerleri ile sırasıyla, doğruluk metriği için %7,79, kesinlik metriği için %7,60, duyarlılık metriği için %9,45, F1-skor metriği için %8,41 artış sağlanmıştır. Elde edilen bu performans artışı ironi ve iğneleme tespitinde bağlam geçmişin sınıflandırma üzerindeki pozitif etkisini göstermektedir. Veri kümesi mimarisi olarak en temel olan bir-gram veri kümesinden tüm-gram veri kümesine kadar en başarılı olan model açısından düzenli bir performans artışı gözlemlenmektedir. En başarılı iki model olan BERTurk ve ConvBERTurk modellerinin veri kümelerindeki elde ettiği metrik değerleri Çizelge 4.12'de sunulmuştur. BERTurk modeli sadece üç-gram veri kümesinden dört-gram veri kümesine geçişte performans artışı sağlayamamış olmasına rağmen bağlamın genişlemesiyle metriklerde artış gözlemlenmiştir. Model tüm-gram veri kümesinde elde ettiği metrik değerleri ile bir-gram veri kümesine göre sırasıyla doğruluk metriği için %7,09, kesinlik metriği için %8,16, duyarlılık metriği için %8,90, F1-skor metriği için %8,50 artış sağladığı gözlemlenmektedir. ConvBERTurk modeli bir-gram ve tüm-gram veri kümeleri arasında bağlamın genişlemesi ile her adımda performans artışı sağlamıştır. tüm-gram veri kümesinde elde ettiği metrik değerleri ile bir-gram veri kümesine göre sırasıyla doğruluk metriği için %16,19, kesinlik metriği için %14,16,

duyarlılık metriği için %16,41, F1-skor metriği için %15,22 artış sağladığı gözlemlenmektedir.

Çizelge 4.12. Veri kümelerinde en başarılı modellerin performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
BERTurk	bir-gram	65,63	63,19	61,89	62,58
BERTurk	iki-gram	66,72	64,72	63,50	64,10
BERTurk	üç-gram	67,42	64,73	64,32	64,52
BERTurk	dört-gram	67,07	65,08	63,99	64,50
BERTurk	tüm-gram	**70,28**	**68,35**	**67,40**	**67,90**
ConvBERTurk	bir-gram	60,89	59,56	58,19	58,88
ConvBERTurk	iki-gram	65,30	64,39	62,00	63,17
ConvBERTurk	üç-gram	65,62	61,58	61,97	61,77
ConvBERTurk	dört-gram	68,76	66,29	65,19	65,75
ConvBERTurk	tüm-gram	**70,74**	**67,99**	**67,74**	**67,84**

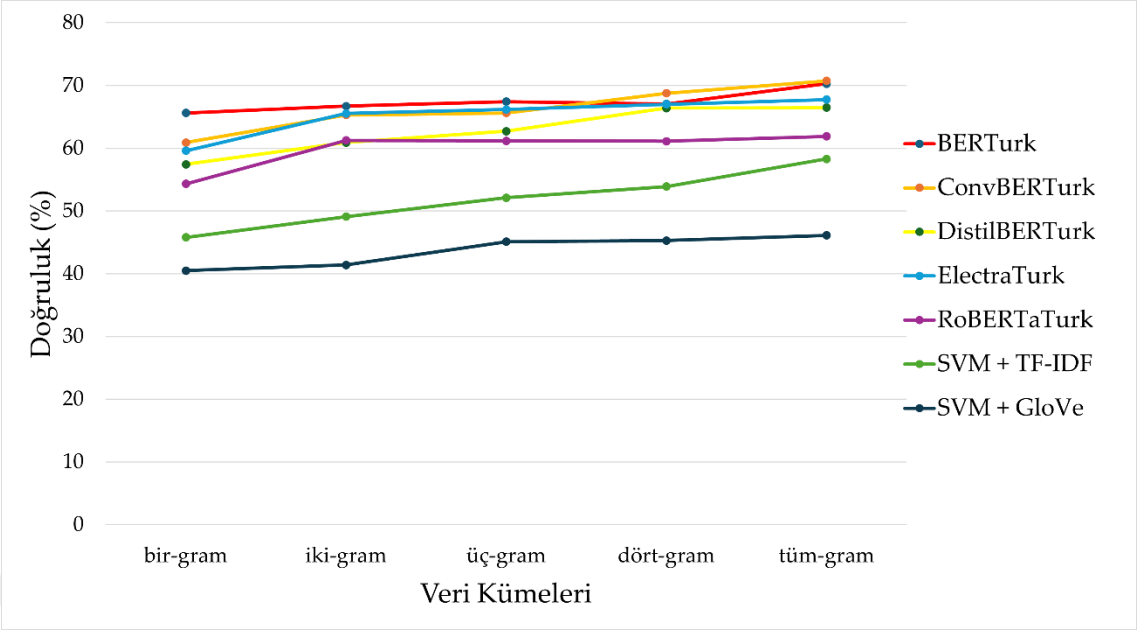
DistilBERTurk, ELECTRATurk, RoBERTaTurk ve SVM modellerinin bağlama göre performansları 4.13'te sunulmuştur. DistilBERTurk modeli bir-gram veri kümesinde %57,45 doğruluk metriği elde ederken tüm-gram veri kümesinde %66,47 doğruluk değeri elde ederek toplamda %15,56 performans artışı gözlemlenmiştir. ELECTRATurk modeli bir-gram veri kümesinde %59,59 doğruluk metriği elde ederken tüm-gram veri kümesinde %67,76 doğruluk değeri elde ederek toplamda %13,67 performans artışı gözlemlenmiştir. RoBERTaTurk modeli bir-gram veri kümesinde %54,32 doğruluk metriği elde ederken tüm-gram veri kümesinde %61,88 doğruluk değeri elde ederek toplamda %14,02 performans artışı gözlemlenmiştir. SVM modelinde TF-IDF kelime gömme yöntemi kullanılarak bir-gram veri kümesinde %45,80 doğruluk metriği elde ederken tüm-gram veri kümesinde %58,30 doğruluk değeri elde ederek toplamda %27,29 performans artışı gözlemlenmiştir. SVM modelinde GloVe kelime gömme yöntemi kullanılarak bir-gram veri kümesinde %40,50 doğruluk metriği elde ederken tüm-gram veri kümesinde %46,10 doğruluk değeri elde ederek toplamda %13,83 performans artışı gözlemlenmiştir. Tüm modellerde kullanılan veri kümesinin bağlam geçmişi arttıkça modellerin performanslarının arttığı gözlemlenmektedir. Modellerin başlangıç performansının düşük olduğu durumlarda en geniş bağlama sahip veri

kümesinde, başlangıç performansı yüksek olan modellere göre daha yüksek bir performans artışı gözlemlenmektedir. En başarılı modeller olan BERTurk ve ConvBERTurk modellerinin bir-gram ve tüm-gram veri kümeleri arasındaki artış yüzdesi düşük kalırken, en düşük performans sahip SVM modellerindeki artış %27,29'a kadar ulaşmaktadır. Elde edilen sonuçlar bağlam geçmişinin modellerin performansında önemli bir artışa neden olduğunu göstermektedir.

Çizelge 4.13. Veri kümelerinde düşük performanslı modellerin sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
DistilBERTurk	bir-gram	57,45	55,61	54,89	55,22
DistilBERTurk	iki-gram	60,91	57,81	57,52	57,66
DistilBERTurk	üç-gram	60,91	57,81	57,52	57,66
DistilBERTurk	dört-gram	66,39	62,70	62,40	62,57
DistilBERTurk	tüm-gram	**66,47**	**63,22**	**63,20**	**63,22**
ELECTRATurk	bir-gram	59,59	55,30	56,56	55,87
ELECTRATurk	iki-gram	65,53	64,08	61,93	62,99
ELECTRATurk	üç-gram	65,53	64,08	61,93	62,99
ELECTRATurk	dört-gram	67,00	62,22	63,72	62,93
ELECTRATurk	tüm-gram	**67,76**	**65,05**	**64,87**	**64,96**
RoBERTaTurk	bir-gram	54,32	50,93	50,72	50,80
RoBERTaTurk	iki-gram	**61,24**	**60,99**	**56,83**	**58,84**
RoBERTaTurk	üç-gram	61,16	56,84	56,61	56,72
RoBERTaTurk	dört-gram	61,11	57,62	56,51	57,05
RoBERTaTurk	tüm-gram	**61,88**	52,00	57,03	54,39
SVM + TF-IDF	bir-gram	45,80	44,20	45,80	44,98
SVM + TF-IDF	iki-gram	49,10	46,80	49,10	47,93
SVM + TF-IDF	üç-gram	49,10	46,80	49,10	47,93
SVM + TF-IDF	dört-gram	53,90	53,10	53,90	53,50
SVM + TF-IDF	tüm-gram	**58,30**	**57,80**	**58,30**	**58,05**
SVM + GloVe	bir-gram	40,50	38,50	40,50	39,47
SVM + GloVe	iki-gram	41,40	41,60	41,40	41,50
SVM + GloVe	üç-gram	41,40	41,60	41,40	41,50
SVM + GloVe	dört-gram	45,30	44,30	45,30	44,80
SVM + GloVe	tüm-gram	**46,10**	**45,90**	**46,10**	**46,00**

Modellerin veri kümelerinin bağlamının genişlemesine göre gösterdikleri performans durumu Şekil 4.1'de sunulmuştur.



Şekil 4.1. Modellerin performans gelişim eğrisi

4.2. Kolektif Öğrenmenin Performans Değerlendirmesi

Çalışmanın bu bölümünde modellerin bireysel performansları ile önerilen kolektif öğrenme yönteminin performansının veri kümelerine göre karşılaştırılması yapılacaktır. Eğitim ve test aşamalarında 80:20 oranında eğitim-test veri kümeleri kullanılmıştır. BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk, RoBERTaTurk ve SVM yöntemlerinin bir-gram, iki-gram, üç-gram, dört-gram ve tüm-gram veri kümeleri üzerinde bireysel ve önerilen kolektif öğrenme yönteminin sınıflandırma performansına ait sonuçlar değerlendirilmiştir. Modellerin değerlendirmesinde elde edilen sonuçlar doğruluk, kesinlik, duyarlılık ve F1-skor performans metrikleri ile değerlendirilmiştir. Modellerin veri kümelerine göre performans sonuçları Çizelge 4.14-4.19 arasında sunulmuştur. Ayrıca kolektif yöntemin veri kümeleri üzerinde elde ettiği performanslar analiz edilerek değerlendirilmiştir. Kolektif yöntem hem bireysel ince ayarlanmış modellerle hem de bu modellerin performans ortalamalarına göre kıyaslanmıştır.

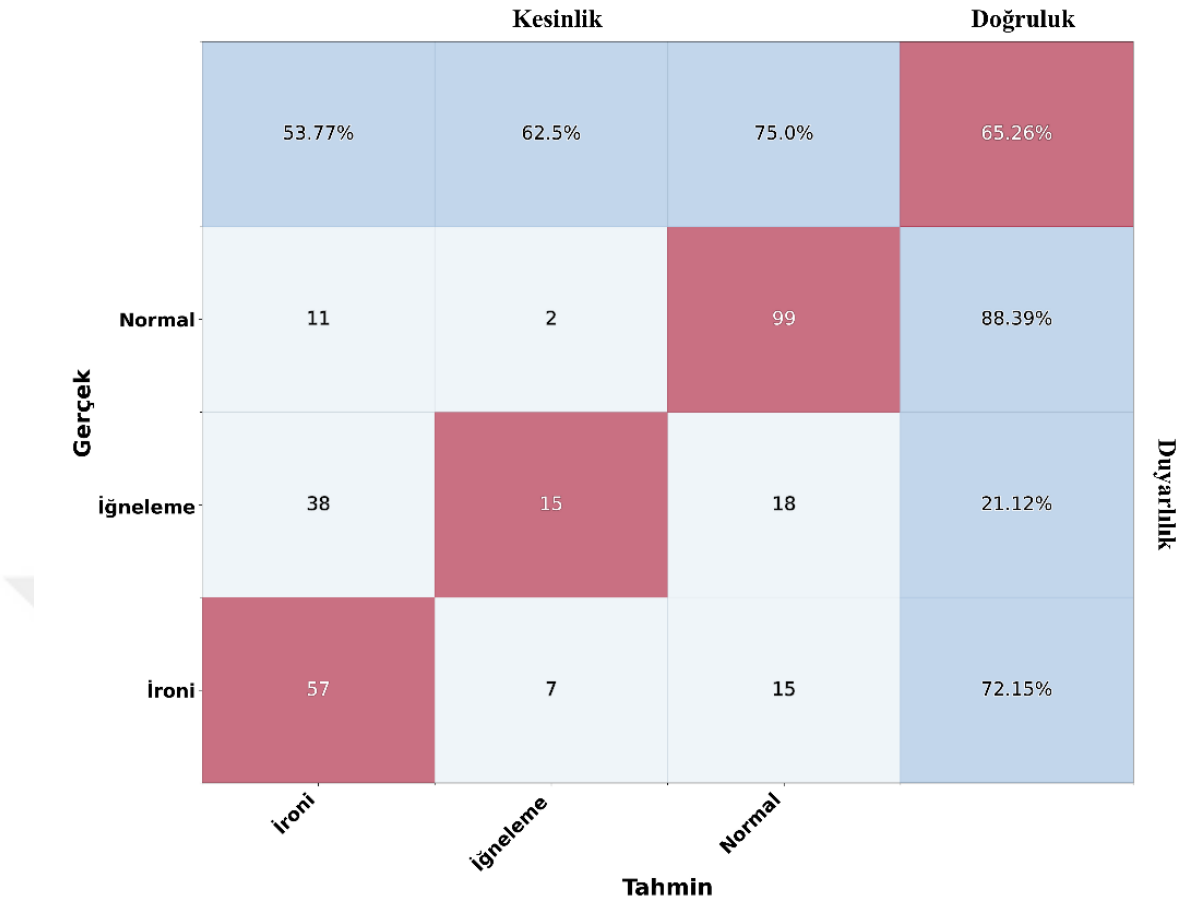
Çizelge 4.14'te bir-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda kolektif öğrenme modeli %65,25 doğruluk, %63,75 kesinlik, %60,55 duyarlılık ve %62,12 F1-skor metrik değerleri ile diğer modellerden daha başarılı olurken ikinci başarılı olan DistilBERTurk modelinden, sırasıyla doğruluk

metriği için %6,19, kesinlik metriği için %18,97, duyarlılık metriği için %16,87, F1-skor metriği için %18,38 artış sağladığı gözlemlenmektedir. Dönüştürücü temelli modellerin performans metrik ortalamaları incelendiğinde, ortalama değerlere göre kolektif öğrenme yöntemi doğruluk metriğinde %11,86, kesinlik metriğinde %20,87, duyarlılık metriğinde %12,27 ve F1-skor metriğinde %17,54 oranında daha iyi sonuçlar elde etmiştir.

Çizelge 4.14. Modellerin bir-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)	Ağırlık
BERTurk	bir-gram	58,02	54,55	54,62	54,58	0,2916
ConvBERTurk	bir-gram	54,20	60,31	57,97	59,12	0,0671
DistilBERTurk	bir-gram	61,45	53,58	51,80	52,48	0,0086
ELECTRATurk	bir-gram	56,49	40,83	55,62	47,02	0,6122
RoBERTaTurk	bir-gram	61,45	52,48	49,71	51,06	0,0205
Ortalama	bir-gram	58,32	52,75	53,94	52,85	-
Kolektif Öğrenme	bir-gram	**65,25**	**63,75**	**60,55**	**62,12**	-

Kolektif öğrenmesine ait karmaşıklık matrisi Şekil 4.2'de sunulmuştur. Kolektif öğrenme yönteminde kullanılan ARO algoritması için tavşan sayısı 50-1000 arasında arama yapılarak 50 tavşan ve 50 yineleme kullanılmıştır.



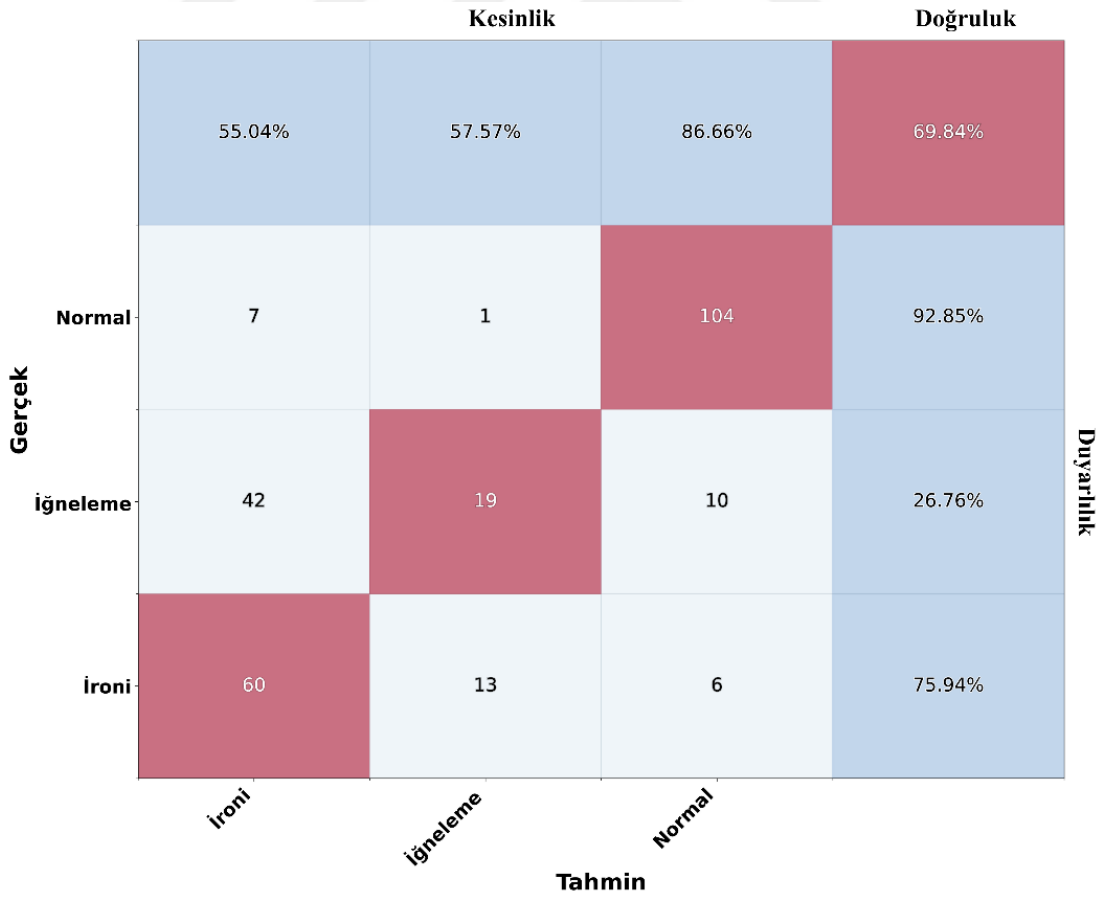
Şekil 4.2. Kolektif öğrenme bir-gram veri kümesi karmaşıklık matrisi

Çizelge 4.15'te iki-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda kolektif öğrenme modeli %69,84 doğruluk, %66,42 kesinlik, %65,18 duyarlılık ve %65,80 F1-skor metrik değerleri ile diğer modellerden daha başarılı olurken ikinci başarılı olan ELECTRATurk modeli, sırasıyla doğruluk metriği için %7,64, kesinlik metriği için %12,25, duyarlılık metriği için %9,18, F1-skor metriği için %10,72 artış sağladığı gözlemlenmektedir. Dönüştürücü temelli modellerin performans metrik ortalamaları incelendiğinde, ortalama değerlere göre kolektif öğrenme yöntemi doğruluk metriğinde %15,55, kesinlik metriğinde %16,03, duyarlılık metriğinde %15,38 ve F1-skor metriğinde %15,74 oranında daha iyi sonuçlar elde etmiştir.

Çizelge 4.15. Modellerin iki-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)	Ağırlık
BERTurk	iki-gram	61,07	56,26	56,50	56,38	0,0525
ConvBERTurk	iki-gram	62,21	59,26	58,45	58,85	0,0369
DistilBERTurk	iki-gram	57,63	52,68	53,22	52,95	0,0601
ELECTRATurk	iki-gram	64,89	59,17	59,70	59,43	0,4594
RoBERTaTurk	iki-gram	56,49	58,91	54,60	56,68	0,3912
Ortalama	iki-gram	60,45	57,25	56,49	56,85	-
Kolektif Öğrenme	iki-gram	**69,84**	**66,42**	**65,18**	**65,80**	-

Kolektif öğrenmeye ait karmaşıklık matrisi Şekil 4.3'te sunulmuştur. Kolektif öğrenme yönteminde kullanılan ARO algoritması için 130 tavşan ve 50 yineleme kullanılmıştır.



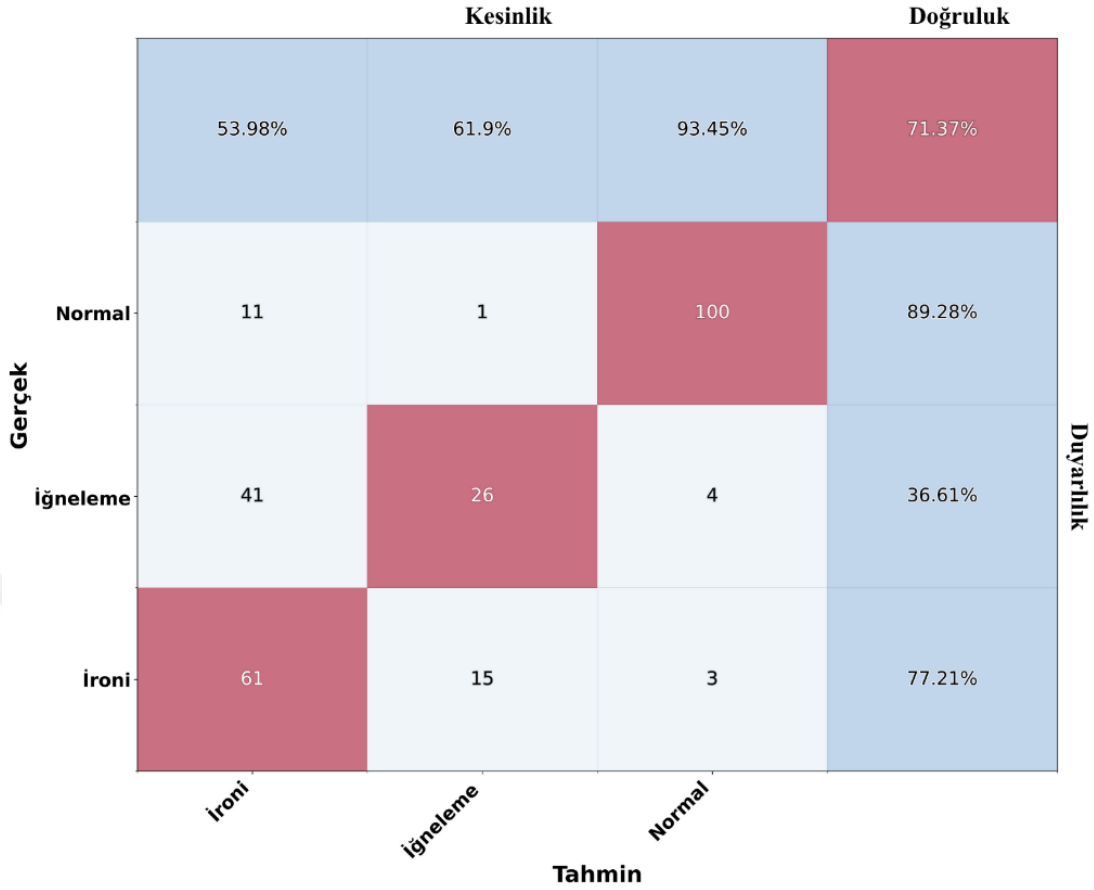
Şekil 4.3. Kolektif öğrenme iki-gram veri kümesi karmaşıklık matrisi

Çizelge 4.16'da üç-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda kolektif öğrenme modeli %71,37 doğruluk, %69,78 kesinlik, %67,70 duyarlılık ve %68,73 F1-skor metrik değerleri ile diğer modellerden daha başarılı olurken ikinci başarılı olan ELECTRATurk modelinden, sırasıyla doğruluk metriği için %7,47, kesinlik metriği için %10,70, duyarlılık metriği için %8,05, F1-skor metriği için %9,36 artış sağladığı gözlemlenmektedir. Dönüştürücü temelli modellerin performans metrik ortalamaları incelendiğinde, ortalama değerlere göre kolektif öğrenme yöntemi doğruluk metriğinde %12,25, kesinlik metriğinde %12,84, duyarlılık metriğinde %11,79 ve F1-skor metriğinde %12,32 oranında daha iyi sonuçlar elde etmiştir.

Çizelge 4.16. Modellerin üç-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)	Ağırlık
BERTurk	üç-gram	62,50	61,68	60,44	61,06	0,0434
ConvBERTurk	üç-gram	65,27	61,51	61,12	61,31	0,0761
DistilBERTurk	üç-gram	64,50	64,32	61,58	62,93	0,4505
ELECTRATurk	üç-gram	66,41	63,04	62,66	62,85	0,4267
RoBERTaTurk	üç-gram	59,24	58,65	57,03	57,83	0,0032
Ortalama	üç-gram	63,58	61,84	60,56	61,19	-
Kolektif Öğrenme	üç-gram	**71,37**	**69,78**	**67,70**	**68,73**	-

Kolektif öğrenmeye ait karmaşıklık matrisi Şekil 4.4'te sunulmuştur. Kolektif öğrenme yönteminde kullanılan ARO algoritması için 190 tavşan ve 50 yineleme kullanılmıştır.



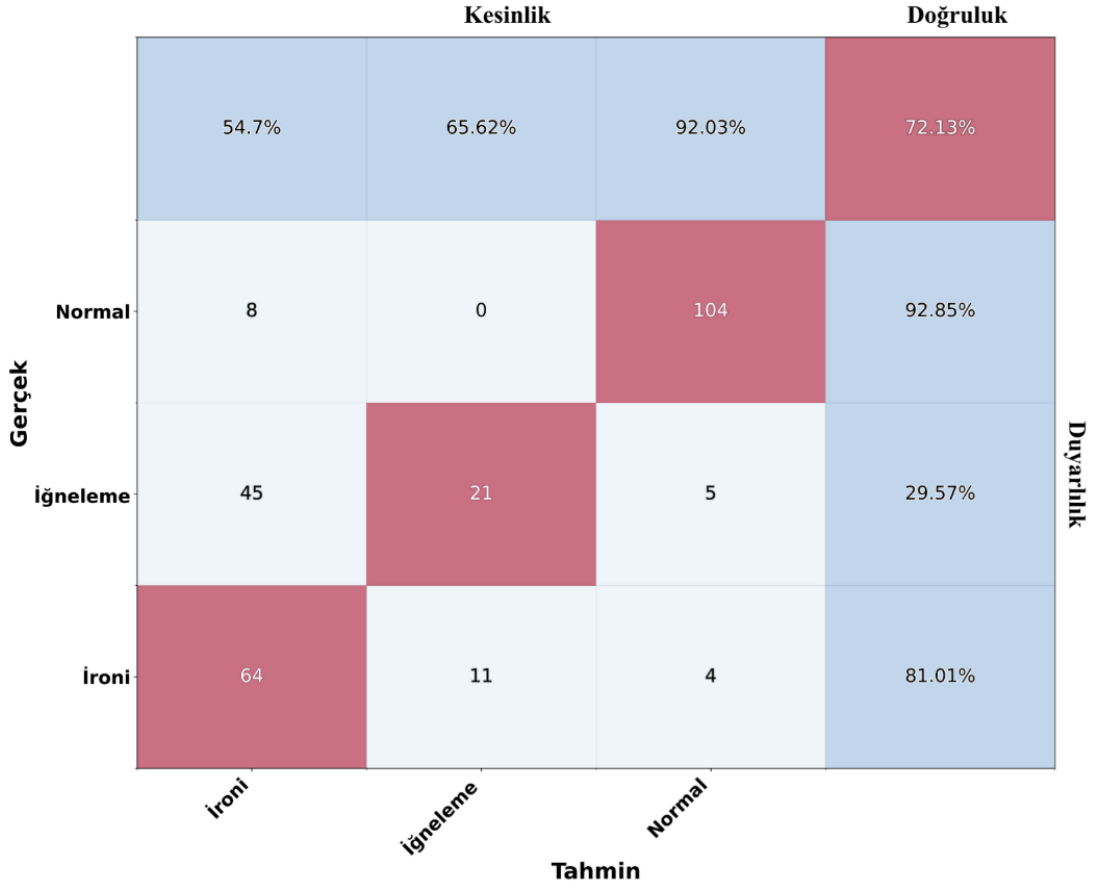
Şekil 4.4. Kolektif öğrenme üç-gram veri kümesi karmaşıklık matrisi

Çizelge 4.17'de dört-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda kolektif öğrenme modeli %72,13 doğruluk, %70,78 kesinlik, %67,81 duyarlılık ve %69,27 F1-skor metrik değerleri ile diğer modellerden daha başarılı olurken ikinci başarılı olan ELECTRATurk modelinden, sırasıyla doğruluk metriği için %8, kesinlik metriği için %18,41, duyarlılık metriği için %12,45, F1-skor metriği için %15,39 artış sağladığı gözlemlenmektedir. Dönüştürücü temelli modellerin performans metrik ortalamaları incelendiğinde, ortalama değerlere göre kolektif öğrenme yöntemi doğruluk metriğinde %11,31, kesinlik metriğinde %17,15, duyarlılık metriğinde %12,49 ve F1-skor metriğinde %14,78 oranında daha iyi sonuçlar elde etmiştir.

Çizelge 4.17. Modellerin dört-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları

Model	Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)	Ağırlık
BERTurk	dört-gram	64,12	59,90	59,60	59,75	0,1992
ConvBERTurk	dört-gram	61,83	62,20	61,70	61,95	0,2988
DistilBERTurk	dört-gram	66,03	62,95	63,00	62,98	0,3102
ELECTRATurk	dört-gram	66,79	59,77	60,30	60,03	0,0991
RoBERTaTurk	dört-gram	65,27	57,29	56,80	57,04	0,0927
Ortalama	dört-gram	64,80	60,42	60,28	60,35	-
Kolektif Öğrenme	dört-gram	**72,13**	**70,78**	**67,81**	**69,27**	-

Kolektif öğrenmeye ait karmaşıklık matrisi Şekil 4.5'te sunulmuştur. Kolektif öğrenme yönteminde kullanılan ARO algoritması için 610 tavşan ve 50 yineleme kullanılmıştır.



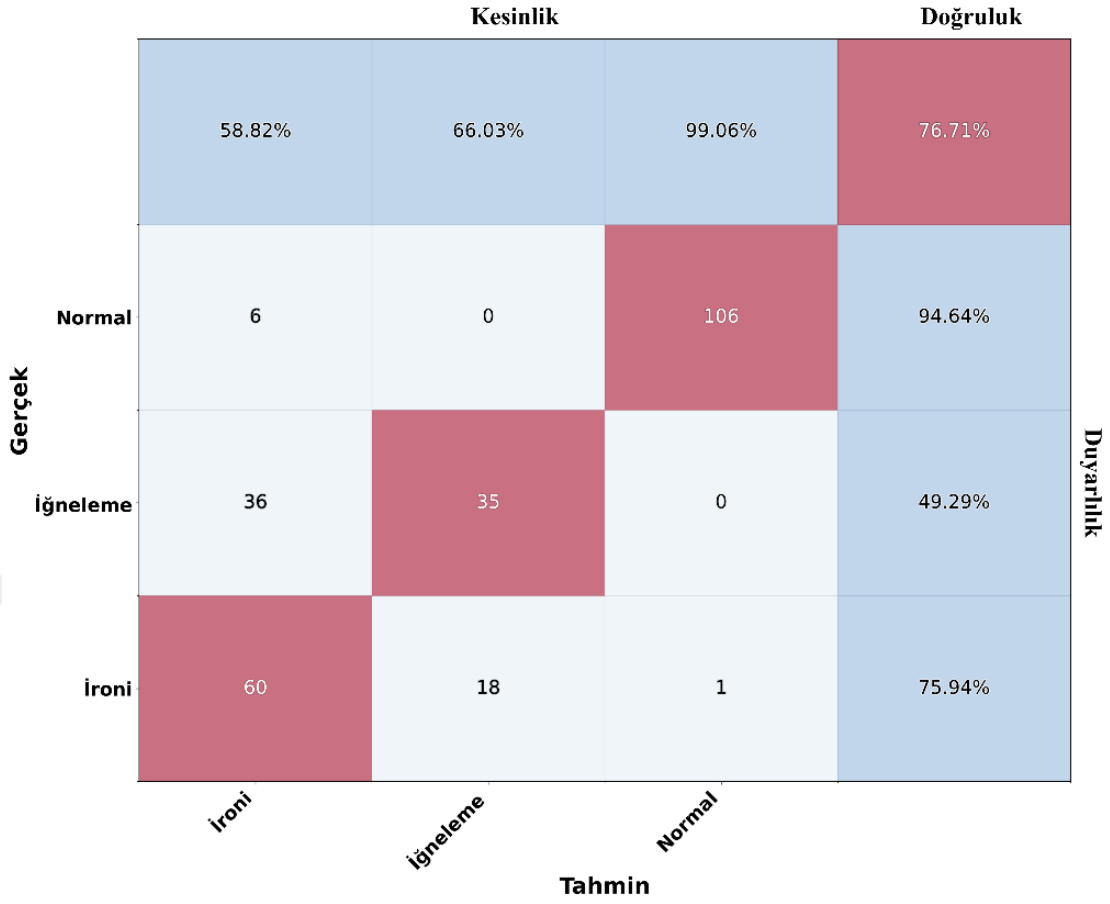
Şekil 4.5. Kolektif öğrenme dört-gram veri kümesi karmaşıklık matrisi

Çizelge 4.18'de tüm-gram veri kümesinde modellerin elde ettiği sonuçlar sunulmuş olup, elde edilen sonuçlar doğrultusunda kolektif öğrenme modeli %76,71 doğruluk, %74,64 kesinlik, %73,29 duyarlılık ve %73,96 F1-skor metrik değerleri ile diğer modellerden daha başarılı olurken ikinci başarılı olan ConvBERTurk modelinden, sırasıyla doğruluk metriği için %6,90, kesinlik metriği için %9,25, duyarlılık metriği için %7,29, F1-skor metriği için %8,25 artış sağladığı gözlemlenmektedir. Ayrıca 4.18'de dönüştürücü temelli modellerin performans metrik ortalamaları verilmiştir. Ortalama değerlere göre kolektif öğrenme yöntemi doğruluk metriğinde %13,93, kesinlik metriğinde %18,14, duyarlılık metriğinde %16,96 ve F1-skor metriğinde %17,56 oranında daha iyi sonuçlar elde etmiştir.

Çizelge 4.18. Modellerin tüm-gram veri kümesinde bireysel ve kolektif öğrenme performans sonuçları

Model	**Veri Kümesi**	**Doğruluk (%)**	**Kesinlik (%)**	**Duyarlılık (%)**	**F1-Skor (%)**	**Ağırlık**
BERTurk	tüm-gram	69,85	66,98	66,31	66,64	0,0639
ConvBERTurk	tüm-gram	71,76	68,32	68,31	68,32	0,4383
DistilBERTurk	tüm-gram	66,41	61,10	60,64	60,87	0,0346
ELECTRATurk	tüm-gram	70,23	69,24	65,76	67,46	0,4601
RoBERTaTurk	tüm-gram	58,40	50,29	52,29	51,27	0,0031
Ortalama	tüm-gram	67,33	63,18	62,66	62,91	-
Kolektif Öğrenme	tüm-gram	**76,71**	**74,64**	**73,29**	**73,96**	-

Kolektif öğrenmeye ait karmaşıklık matrisi Şekil 4.6'da sunulmuştur. Kolektif öğrenme yönteminde kullanılan ARO algoritması için 390 tavşan ve 50 yineleme kullanılmıştır.



Şekil 4.6. Kolektif öğrenme tüm-gram veri kümesi karmaşıklık matrisi

Kolektif öğrenme yönteminin tüm veri kümelerinde elde ettiği performans metrikleri 4.19'da sunulmuştur. Tablo incelendiğinde kolektif öğrenme yöntemi bir-gram veri kümesinde %66,41 doğruluk metriği elde ederken bağlam genişliği diğer veri kümelerinde artarken elde edilen doğruluk performansının da arttığı görülmektedir. Artan performans tüm-gram yani bağlamın en geniş olduğu veri kümesinde %76,71 doğruluk ile en yüksek değere ulaşmaktadır. Aynı şekilde diğer metriklerde de bir-gram ve tüm-gram veri kümelerindeki sonuçlar incelendiğinde iki veri kümesi arasında performans artışı görülmektedir.

Çizelge 4.19. Kolektif öğrenme yönteminin veri kümesi türüne göre performans sonuçları

Veri Kümesi	Tavşan Sayısı	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
bir-gram	50	65,25	63,75	60,55	62,12
iki-gram	130	69,84	66,42	65,18	65,80
üç-gram	190	71,37	69,78	67,70	68,73
dört-gram	610	72,13	70,78	67,81	69,27
tüm-gram	390	**76,71**	**74,64**	**73,29**	**73,96**

Kolektif öğrenme için kullanılan yapay tavşan optimizasyon algoritmasının 50 tekrar edilerek modellere ait örnek ağırlıkların istatistikleri Çizelge 4.20'de sunulmaktadır. Dönüştürücü temelli yöntemlerin eğitimi tamamlandıktan sonra, ağırlıklarının belirlenmesi için optimizasyon algoritması tarafından ağırlıklarının belirlenmesi için iterasyon sayısı kadar çalıştırılıp en iyi çözüm adayı bulma işlemi gerçekleştirilmektedir. En iyi çözüm adayı belirlendikten sonra bu işlem 50 kez tekrar edilerek, bireysel modellere atanan en iyi 50 ağırlığın istatistikleri çizelgede verilmektedir. Bir-gram, iki-gram, üç-gram ve tüm-gram veri kümelerinde en yüksek ortalama değerler ELECTRATurk modeli için belirlenirken, tüm-gram veri kümesinde ConvBERTurk modeline atanmıştır. Ağırlık verileri incelendiğinde, optimizasyon algoritması tarafından hesaplanan ortalama değerlerin ve standart sapmaların istatistiksel açıdan beklenen sınırlar dahilinde olduğu görülmektedir. Özellikle ortalama ve ortanca değerlerinin birbirine yakınlığı, algoritma tarafından kolektif öğrenme için belirlenen çözüm adayının ağırlık örneklerinin tutarlı olduğunu göstermektedir.

Çizelge 4.20. Optimizasyon algoritmasının ürettiği ağırlıkların istatistikleri

Model	Veri Kümesi	Minimum	Maksimum	Ortalama	Ortanca	Standart Sapma
BERTurk	bir-gram	0.2479	0.3525	0.2969	0.3004	0.0198
	iki-gram	0.0000	0.0537	0.0120	0.0047	0.0145
	üç-gram	0.0000	0.0445	0.0178	0.0154	0.0138
	dört-gram	0.1026	0.2288	0.2109	0.2144	0.0225
	tüm-gram	0.0000	0.0804	0.0057	0.0000	0.0129
ConvBERTurk	bir-gram	0.0000	0.0942	0.0217	0.0011	0.0282
	iki-gram	0.0000	0.0523	0.0173	0.0112	0.0184
	üç-gram	0.0000	0.1367	0.0353	0.0139	0.0456
	dört-gram	0.2519	0.5003	0.4086	0.4736	0.0953
	tüm-gram	0.3970	0.4585	0.4406	0.4486	0.0175
DistilBERTurk	bir-gram	0.0000	0.0172	0.0033	0.0000	0.0048
	iki-gram	0.0096	0.0682	0.0332	0.0347	0.0094
	üç-gram	0.1295	0.4702	0.3440	0.4357	0.1305
	dört-gram	0.1527	0.4982	0.2241	0.2001	0.0692
	tüm-gram	0.0339	0.1275	0.0946	0.1005	0.0198
ELECTRATurk	bir-gram	0.3656	0.6213	0.5634	0.6008	0.0690
	iki-gram	0.4410	0.5087	0.4903	0.4999	0.0217
	üç-gram	0.3886	0.4333	0.4120	0.4145	0.0111
	dört-gram	0.0089	0.1712	0.1533	0.1612	0.0287
	tüm-gram	0.4508	0.5284	0.4631	0.4625	0.0109
RoBERTaTurk	bir-gram	0.0000	0.0354	0.0068	0.0000	0.0109
	iki-gram	0.1239	0.4026	0.3209	0.3109	0.0737
	üç-gram	0.0000	0.0044	0.0006	0.0000	0.0013
	dört-gram	0.0640	0.1050	0.0838	0.0824	0.0110
	tüm-gram	0.0000	0.0544	0.0378	0.0379	0.0094

Veri etiketleme sürecinde kullanılan SarcaTurkGPT isimli aracın veri kümeleri üzerinde elde ettiği metrikler Çizelge 4.21'de sunulmaktadır.

Çizelge 4.21. Veri kümeleri üzerinde SarcaTurkGPT'nin elde ettiği performans metrikleri

Veri Kümesi	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1-Skor (%)
bir-gram	**59,16**	**58,90**	**57,69**	**58,29**
iki-gram	59,16	56,01	55,99	56,00
üç-gram	53,82	53,39	52,17	52,27
dört-gram	53,05	49,21	49,84	49,52
tüm-gram	52,67	53,86	51,12	52,45

Bağlamsal veri kümelerinde %20 test veri seti kullanılarak gerçekleştirilen SarcaTurkGPT ile gerçekleştirilen tahmin performansında, bir-gram veri kümesinde %59,16 doğruluk, %58,90 kesinlik, %57,69 duyarlılık ve %58,29 F1-skor metrikleri ile modelin en başarılı olduğu veri kümesi olmuştur. Model tüm-gram veri kümesinde %52,67 doğruluk değeri ile son sırada yer almaktadır. Veri kümelerinin bağlamı genişledikçe modelin tahmin performansında düşüşler meydana gelmiştir. Bağlama duyarlı geliştirilen bireysel modeller ve kolektif model, veri kümelerinin bağlamı arttıkça performansını artırırken, bağlam eğitimi olmadan tahminler gerçekleştiren SarcaTurkGPT'nin performansı bu durumdan olumsuz etkilenmiştir.

5. TARTIŞMA ve SONUÇ

Çalışmada, Türkçe metinlerde ironik ve iğneleyici ifadelerin tespit edilmesinde beş farklı önceden eğitilmiş dönüştürücü tabanlı modellerden BERTurk, ConvBERTurk, DistilBERTurk, ELECTRATurk ve RoBERTaTurk'ün performansları değerlendirilmiş ve bu modeller aynı metin üzerinden özel olarak hazırlanmış bir-gram, iki-gram ve üç-gram, dört-gram ve tüm-gram veri kümeleri üzerinde ince ayar yapılarak test edilmiştir. Deneysel süreçte, sınıflandırma başarısını artırmak amacıyla bireysel modellerin tahminini birleştiren ARO algoritması tabanlı bir kolektif öğrenme yaklaşımı da uygulanmıştır.

Modellerin çapraz doğrulama tekniğine göre veri kümeleri üzerinde elde ettiğini sonuçlar incelendiğinde bağlamı en dar olan bir-gram veri kümesinden, bağlamı en geniş olan tüm-gram veri kümesi arasında doğruluk, kesinlik, duyarlılık ve F1-skor metriklerinde sırasıyla %65,63, %63,19, %61,89, %62,58'den %70,74, %67,99, %67,74, %67,84'e arttığı görülmektedir. bir-gram, iki-gram ve üç-gram veri kümelerinde BERTurk, dört-gram ve tüm-gram veri kümelerinde ise ConvBERTurk modeli performans açısından önde yer almaktadır. Tüm veri kümelerinde SVM ile TF-IDF ve GloVe kelime temsil yöntemlerinin birlikte kullanıldığı yöntem son sırada yer almaktadır. BERTurk modeli bir-gram veri kümesinde %65,63 doğruluk elde ederken, tüm-gram veri kümesinde %70,28 elde etmiştir. Benzer şekilde doğruluk metriğinde ConvBERTurk modeli %60,89'dan %70,74'e, DistilBERTurk modeli %57,45'ten %66,47'ye, ELECTRATurk modeli %59,59'dan %67,76'ya, RoBERTaTurk modeli %54,32'den %61,66'e, SVM+TF-IDF modeli %45,80'den %58,30'a ve SVM+GloVe modeli %40,50'den %46,10'a artan performans göstermiştir. Modeller genellikle iki-gram veri kümesinden üç-gram veri kümesine ve dört-gram veri kümesine bağlamı genişlediğinde sınırlı performans artışı veya bir önceki veri kümesindeki sonuca benzer sonuçlar elde etmiştir. Son ve en geniş bağlama sahip olan tüm-gram veri kümesinde tüm modeller performans artışı göstermiştir.

Çapraz doğrulama sonucunda elde edilen genel performans metriklerinde olduğu gibi bağlamın sınıf performansı üzerinde artan bir etkiye sahip olduğu görülmüştür. İroni sınıfı için bir-gram veri kümesinde en yüksek doğruluk metriği BERTurk modeli ile

%69, tüm-gram veri kümesinde en yüksek doğruluk metriği %72 ile BERTurk ve ConvBERTurk modelleri ile elde edilmiştir. İğneleme sınıfı için bir-gram veri kümesinde en yüksek doğruluk metriği BERTurk modeli ile %72, tüm-gram veri kümesinde en yüksek doğruluk metriği %74 ile ConvBERTurk modeli ile elde edilmiştir. Normal sınıfı için bir-gram veri kümesinde en yüksek doğruluk metriği BERTurk modeli ile %89, tüm-gram veri kümesinde en yüksek doğruluk metriği %97 ile ConvBERTurk modeli ile elde edilmiştir.

Çalışmada kullanılan klasik makine öğrenmesi algoritması olan SVM, TF-IDF ve GloVe kelime temsil yöntemlerinin birlikte kullanıldığı yöntemlerde veri kümelerinin bağlam genişliğinin artmasıyla orantılı olarak performans metriklerine artış gözlemlenmiştir. TF-IDF yöntemi, GloVe yöntemine göre tüm metriklerde ve tüm veri kümelerinde daha iyi sonuçlar elde etmiştir. SVM yöntemlerinin elde ettiği artış oranı dönüştürücü tabanlı modellere benzer olsa da tüm veri kümelerinde son sırada yer almaktadır. Bu durum dönüştürücü tabanlı modellerin hesaplama maliyeti ve model boyutlarının, klasik makine öğrenmesi yöntemlerine göre daha fazla olmasına rağmen başarı performansı ölçütlerinde önemli bir fark ortaya koymuştur.

Dönüştürücü tabanlı modellerin veri kümeleri üzerinde 80:20 oranında eğitim ve test bölünmesine göre elde ettiği sonuçlar incelendiğinde, bireysel modellerin veri kümelerindeki bağlamın artmasıyla başarı metriklerinde artış görülmektedir. Modellerin performans artışı, modelleri kullanarak nihai tahmin yapan kolektif öğrenme yönteminin artışına yansıdığı görülmektedir. Bireysel modellerin başarısının artmasıyla birlikte kolektif öğrenme yöntemi, bağlamı en dar bir-gram veri kümesinde %65,25 doğruluk metriği performansı elde ederken bağlamın en geniş olduğu tüm-gram veri kümesinde %76,71 doğruluk metriği elde etmiştir. Elde edilen sonuçlar kolektif öğrenme yönteminin bireysel modellerin sonuçlarını başarılı ve etkin şekilde birleştirerek tüm veri kümeleri üzerinde bireysel modellere göre daha yüksek performans elde etmiştir.

Bireysel model performanslarını geliştiren kolektif öğrenme yöntemi, tüm modellerin eğitilmesi ve performans ölçümünün yapılması nedeniyle hesaplama, zaman ve

donanım kaynağı gereksinimleri nedeniyle gerçek zamanlı uygulamalar açısından kullanımında sorun teşkil edebilmektedir.

Genel olarak, bu çalışma hem kullanılan model türünün hem de veri kümesi mimarisinin metin sınıflandırma performansı üzerinde belirleyici olduğunu ve bu iki durumun ARO gibi optimizasyon yöntemleriyle desteklendiğinde, başarıyı anlamlı düzeyde artırabileceğini ortaya koymaktadır. Elde edilen sonuçlar, Türkçe gibi düşük kaynaklı dillerde derin öğrenme tabanlı çözümler geliştirirken model seçimi, veri yapısı ve kolektif yöntemlerin birlikte değerlendirilmesi gerektiğine işaret etmektedir.

KAYNAKLAR

Abadeer, M. (2020). Assessment of DistilBERT performance on Named Entity Recognition task for the detection of Protected Health Information and medical concepts. Proceedings of the 3rd Clinical Natural Language Processing Workshop, 158–167. https://doi.org/10.18653/V1/2020.CLINICALNLP-1.18

Abdar, M., Basiri, M. E., Yin, J., Habibnezhad, M., Chi, G., Nemati, S., & Asadi, S. (2020). Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets. Renewable and Sustainable Energy Reviews, 124. https://doi.org/10.1016/j.rser.2020.109781

Abercrombie, G., & Hovy, D. (2016). Putting sarcasm detection into context: The effects of class imbalance and manual labelling on supervised machine classification of twitter conversations. 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Student Research Workshop. https://doi.org/10.18653/v1/ p16-3016

Abulaish, M., Kamal, A., & Zaki, M. J. (2020). A survey of figurative language and its computational detection in online social networks. https://doi.org/10.1145/ 3375547

Akalın, Ş. H. (2009). Türk dili: Dünya dili. Türk Dili, 687, 195–204.

Al-Ghadhban, D., Alnkhilan, E., Tatwany, L., & Alrazgan, M. (2018). Arabic sarcasm detection in Twitter. Proceedings - 2017 International Conference on Engineering and MIS, ICEMIS 2017, 2018-January. https://doi.org/10.1109/ICEMIS.2017. 8272990

Almuqren, L., & Cristea, A. (2021). AraCust: a Saudi Telecom Tweets corpus for sentiment analysis. PeerJ Computer Science, 7, e510.

Amir, S., Wallace, B. C., Lyu, H., Carvalho, P., & Silva, M. J. (2016). Modelling context with user embeddings for sarcasm detection in social media. CoNLL 2016 - 20th SIGNLL Conference on Computational Natural Language Learning, Proceedings. https://doi.org/10.18653/v1/k16-1017

Anan, R., Apon, T. S., Hossain, Z. T., Modhu, E. A., Mondal, S., & Alam, M. D. G. R. (2023). Interpretable Bangla Sarcasm Detection using BERT and Explainable AI. 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC),1272–1278.

Ashwitha, A., Shruthi, G., Shruthi, H. R., Upadhyaya, M., Ray, A. P., & Manjunath, T. C. (2020). Sarcasm detection in natural language processing. Materials Today: Proceedings, 37. https://doi.org/10.1016/j.matpr.2020.09.124

Attardo, S. (2000). Irony as relevant inappropriateness. Journal of Pragmatics, 32. https://doi.org/10.1016/S0378-2166 (99)00070-3

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Bamman, D., & Smith, N. (2015). Contextualized Sarcasm Detection on Twitter. Proceedings of the International AAAI Conference on Web and Social Media, 9(1), 574-577.

Barbieri, F., Saggion, H., & Ronzano, F. (2014). Modelling Sarcasm in Twitter, a Novel Approach. Proceedings of the Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.3115/v1/w14-2609

Ştefan, A. C. B. (2022). Automatic sarcasm detection: Systematic literature review. Information, 13(8).

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. Future Generation Computer Systems, 115. https://doi.org/10.1016/j.future.2020.08.005

Baykara, B., & Güngör, T. (2023). Turkish abstractive text summarization using pretrained sequence-to-sequence models. Natural Language Engineering, 29. https://doi.org/ 10.1017/s1351324922000195

Bettelli, G., & Panzeri, F. (2023). "Irony is easy to understand": The role of emoji in irony detection. Intercultural Pragmatics, 20, 467–493.

Bharti, S. K., Vachha, B., Pradhan, R. K., Babu, K. S., & Jena, S. K. (2016). Sarcastic sentiment detection in tweets streamed in real time: a big data approach. Digital Communications and Networks, 2. https://doi.org/10.1016/j.dcan.2016.06.002

Bharti, S. K., Babu, K. S., & Jena, S. K. (2015). Parsing-based sarcasm sentiment recognition in Twitter data. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015. https://doi.org/10.1145/2808797.2808910

Bharti, S. K., Naidu, R., & Babu, K. S. (2020). Hyperbolic Feature-based Sarcasm Detection in Telugu Conversation Sentences. Journal of Intelligent Systems, 30. https://doi.org/10.1515/jisys-2018-0475

Bharti, S. K., Sathya Babu, K., & Jena, S. K. (2017). Harnessing online news for sarcasm detection in hindi tweets. In International conference on pattern recognition and machine intelligence (pp. 679-686). Cham: Springer International Publishing.

Birba, D. E. (2020). A comparative study of data splitting algorithms for machine learning model selection [Master's thesis, KTH Royal Institute of Technology]. DiVA. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-287194

Bolucu, N., & Can, B. (2020). Sarcasm Target Identification with LSTM Networks. https://doi.org/10.1109/SIU49456.2020.9302321

Bölücü, N., & Can, B. (2025). Semantically-Informed Graph Neural Networks for Irony Detection in Turkish. ACM Transactions on Asian and Low-Resource Language Information Processing, 24(1), 1-20.

Bouazizi, M., & Otsuki, T. (2016). A Pattern-Based Approach for Sarcasm Detection on Twitter. IEEE Access, 4. https://doi.org/10.1109/ACCESS.2016.2594194

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies, 11, 1636.

Burfoot, C., & Baldwin, T. (2009). Automatic satire detection: Are you having a laugh? Proceedings of the ACL-IJCNLP 2009 conference short papers, 161–164.

Buyukbas, E. B., Dogan, A. H., Ozturk, A. U., & Karagoz, P. (2021, September). Explainability in irony detection. In International Conference on Big Data Analytics and Knowledge Discovery (pp. 152-157). Cham: Springer International Publishing.

byunal. (2023). roberta-base-turkish-cased-stance-detection [Yazılım]. Hugging Face. https://huggingface.co/byunal/roberta-base-turkish-uncased-stance-detection

Cai, Y., Cai, H., & Wan, X. (2019). Multi-modal sarcasm detection in twitter with hierarchical fusion model. Proceedings of the 57th annual meeting of the association for computational linguistics, 2506–2515.

Cambria, E., Liu, Q., Decherchi, S., Xing, F., & Kwok, K. (2022). SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis. In Proceedings of the thirteenth language resources and evaluation conference (pp. 3829-3839).

Castro, S., Chiruzzo, L., & Rosá, A. (2018). Overview of the HAHA task: Humor analysis based on human annotation at IberEval 2018. Ibereval@ sepln, 187–194.

Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., & Poria, S. (2019). Towards multimodal sarcasm detection (an obviouslyperfectpaper). arXiv preprint arXiv:1906.01815.

Cekinel, R. F., & Karagoz, P. (2022, July). Text-based causal inference on irony and sarcasm detection. In International Conference on Big Data Analytics and Knowledge Discovery (pp. 31-45). Cham: Springer International Publishing.

Cemek, Y., Cidecio, C., Öztürk, A. U., Çekinel, R. F., & Karagöz, P. (2020, October). Investigating the neural models for Irony detection on Turkish informal texts. In 2020 28th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Chakravarthi, B. R. (2025). Sarcasm detection in Tamil and Malayalam YouTube comments. Social Network Analysis and Mining, 15(1), 72.

Chaudhari, P., & Chandankhede, C. (2018). Literature survey of sarcasm detection. Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017, 2018-January. https://doi.org/10. 1109/WiSPNET.2017.8300120

Chauhan, D. S., Singh, G. V., Arora, A., Ekbal, A., & Bhattacharyya, P. (2022). An emoji-aware multitask framework for multimodal sarcasm detection. Knowledge-Based Systems, 257. https://doi.org/10.1016/j.knosys.2022.109924

Chia, Z. L., Ptaszynski, M., Masui, F., Leliwa, G., & Wroczynski, M. (2021). Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection. Information Processing and Management, 58. https://doi.org/10.1016/j.ipm.2021.102600

Chiruzzo, L., Castro, S., Etcheverry, M., Garat, D., Prada, J. J., & Rosá, A. (2019). Overview of HAHA at IberLEF 2019: Humor analysis based on human annotation. Iberlef@ sepln, 132–144.

Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. 8th International Conference on Learning Representations, ICLR 2020. https://arxiv.org/abs/2003.10555v1

Clavel Quintero, Y., & Arco García, L. (2018). Irony detection based on character language model classifiers. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11047 LNCS.

Clift, R. (1999). Irony in conversation. Language in Society, 28. https://doi.org/10.1017/s0047404599004029

Çağla, A. (2024). Sosyal medyada yapay zekâ ile nefret söylemi tespiti [Doktora tezi, Gazi Üniversitesi]. YÖK Tez Merkezi. https://tez.yok.gov.tr

da Rocha Junqueira, J., da Silva, F., Costa, W., Carvalho, R., Bender, A., Correa, U., & Freitas, L. (2023a). BERTimbau in action: An investigation of its abilities in sentiment analysis, aspect extraction, hate speech detection, and irony detection.

The International FLAIRS Conference Proceedings, 36. https://doi.org/10.33076/flairsproc.2023.v36i1.a036p01

da Rocha Junqueira, J., Junior, C. L., Silva, F. L. V., Côrrea, U. B., & de Freitas, L. A. (2023b). Albertina in action: An investigation of its abilities in aspect extraction, hate speech detection, irony detection, and question-answering. In Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL) (pp. 146–155).

Davidov, D., Tsur, O., & Rappoport, A. (2010). Semi-supervised recognition of sarcastic sentences in twitter and Amazon. CoNLL 2010 - Fourteenth Conference on Computational Natural Language Learning, Proceedings of the Conference.

dbmdz. (2020a). convbert-base-turkish-cased [Yazılım]. Hugging Face. https://huggingface.co/dbmdz/convbert-base-turkish-cased

dbmdz. (2020b). distilbert-base-turkish-cased [Yazılım]. Hugging Face. https://huggingface.co/dbmdz/distilbert-base-turkish-cased

dbmdz. (2020c). electra-base-turkish-cased-discriminator [Yazılım]. Hugging Face. https://huggingface.co/dbmdz/electra-base-turkish-cased-discriminator

Delmonte, R., & Stingo, M. (2016). Detecting satire in Italian political commentaries. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9876 LNCS.

del Pilar Salas-Zárate, M., Paredes-Valverde, M. A., Rodriguez-García, M. Á., Valencia- García, R., & Alor-Hernández, G. (2017). Automatic detection of satire in Twitter: A psycholinguistic-based approach. Knowledge-Based Systems, 128, 20–33. https://doi.org/10.1016/J.KNOSYS.2017.04.009

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 4171– 4186.

Devran, Y. (2016). Söylemlerin Dilinden Suriye Sorunu. Marmara İletişim Dergisi. https://doi.org/10.17829/midr.20162520716

Dimovska, J., Angelovska, M., Gjorgjevikj, D., & Madjarov, G. (2018). Sarcasm and Irony Detection in English Tweets. Communications in Computer and Information Science, 940. https://doi.org/10.1007/978-3-030-00825-3_11

Duran, A. (2025). SarcaTurkGPT [Yazılım]. Erişim tarihi: 20 Ocak 2025, https://chatgpt.com/g/g-UlWt7PIuT-sarcaturkgpt

Dülger, O. (2018). Türkçe metinlerde ironi tespiti. Gazi Üniversitesi, Erişim Tarihi, 3, 2020.

Elçi, E. (2019). The rise of populism in Turkey: a content analysis. Journal of Southeast European and Black Sea, 19. https://doi.org/10.1080/14683857.2019.1656875

Enyedi, Z. (2016). Populist Polarization and Party System Institutionalization: The Role of Party Politics in De-Democratization. Problems of Post-Communism, 63. https://doi.org/10.1080/10758216.2015.1113883

Ercalisun, A. B. (2004). Başlangıcından Yirminci Yüzyıla. Türk Dili Tarihi, Ankara: Akçağ.

Ercilasun, A. B. (2013). Türkçenin Dünya Dilleri Arasındaki Yeri. Dil Araştırmaları, 12, 17–22. https://dergipark.org.tr/en/pub/dilarastirmalari/issue/4726/64774

Eskişar, G. M. K., & Çöltekin, Ç. (2022). Emotions Running High? A Synopsis of the State of Turkish Politics through the ParlaMint Corpus. ParlaCLARIN III 2022 -

Workshop on Creating, Enriching and Using Parliamentary Corpora, Proceedings as part of the 13th Edition of the Language Resources and Evaluation Conference, LREC 2022.

Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian Ensemble Learning. Decision Support Systems, 68. https://doi.org/10.1016/j.dss.2014.10.004

Filatova, E. (2012). Irony and sarcasm: Corpus generation and analysis using crowdsourcing. Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012.

Forslid, E., & Wikén, N. (2015). Automatic irony and sarcasm detection in social media. [Yüksek lisans tezi, Uppsala Universitet]. Uppsala University Publications https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-262242

Freitas, L. A. D., Vanin, A. A., Hogetop, D. N., Bochernitsan, M. N., & Vieira, R. (2014). Pathways for irony detection in tweets. Proceedings of the ACM Symposium on Applied Computing. https://doi.org/10.1145/2554850.2555048

Frenda, S., Patti, V., & Rosso, P. (2023). When Sarcasm Hurts: Irony-Aware Models for Abusive Language Detection. https://doi.org/10.1007/978-3-031-42448-9_4

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. Artificial Intelligence Review, 58, 1–49.

Gargiulo, F., Minutolo, A., Guarasci, R., Damiano, E., Pietro, G. D., Fujita, H., & Esposito, M. (2022). An ELECTRA-Based Model for Neural Coreference Resolution. IEEE Access, 10, 75144–75157. https://doi.org/10.1109/ACCESS.2022.3189956

Gedela, R. T., Meesala, P., Baruah, U., & Soni, B. (2023). Identifying sarcasm using heterogeneous word embeddings: a hybrid and ensemble perspective. Soft Computing. https://doi.org/10.1007/s00500-023-08368-6

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., & Reyes, A. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), 470–478.

Ghosh, A., & Veale, T. (2016). Fracking sarcasm using neural network. Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, 161–169.

Ghosh, D., Fabbri, A. R., & Muresan, S. (2018). Sarcasm analysis using conversation context. Computational Linguistics, 44. https://doi.org/10.1162/coli_a_00336

Ghosh, D., Fabbri, A. R., & Muresan, S. (2017). The role of conversation context for sarcasm detection in online interactions. SIGDIAL 2017 - 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference. https://doi.org/10.18653/v1/w17-5523

Goel, P., Jain, R., Nayyar, A., Singhal, S., & Srivastava, M. (2022). Sarcasm detection using deep learning and ensemble learning. Multimedia Tools and Applications, 81. https://doi.org/10.1007/s11042-022-12930-z

Gong, X., Zhao, Q., Zhang, J., Mao, R., & Xu, R. (2020). The design and construction of a Chinese sarcasm dataset. Proceedings of the twelfth language resources and evaluation conference, 5034–5039.

González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). Identifying sarcasm in Twitter: A closer look. ACL-HLT 2011 - Proceedings of the 49th Annual Meeting

of the Association for Computational Linguistics: Human Language Technologies, 2.

Govindan, V., & Balakrishnan, V. (2022). A machine learning approach in analyzing the effect of hyperboles using negative sentiment tweets for sarcasm detection. Journal of King Saud University - Computer and Information Sciences, 34, 5110–5120. https://doi.org/10.1016/J.JKSUCI.2022.01.008

Gregory, H., Li, S., Mohammadi, P., Tarn, N., Draelos, R., & Rudin, C. (2020). A Transformer Approach to Contextual Sarcasm Detection in Twitter. https://doi.org/10. 18653/v1/2020.figlang-1.37

Grover, V., & Banati, H. (2022). Understanding the Sarcastic Nature of Emojis with SarcOji. Proceedings of the Fifth International Workshop on Emoji Understanding and Applications in Social Media, 29–39.

Grover, V., & Banati, H. (2024). An attention approach to emoji focused sarcasm detection. Helion, 10(17).

Gupta, R., Kumar, J., Agrawal, H., & Kunal. (2020). A Statistical Approach for Sarcasm Detection Using Twitter Data. Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020. https://doi.org/10.1109/ ICICCS48265.2020.9120917

Hamza, M. A., Alshahrani, H. J., Hassan, A. Q., Gaddah, A., Allheeib, N., Alsaif, S. A., Al-Onazi, B. B., & Mohsen, H. (2023). Computational Linguistics with Optimal Deep Belief Network Based Irony Detection in Social Media. Computers, Materials and Continua, 75. https://doi.org/10.32604/cmc.2023.035237

Harared, N., & Nurani, S. (2020). Sarcasm: Mock Politeness in the Big Bang Theory. Elite: English and Literature Journal, 7(2), 186-199.

Hee, C. V., Lefever, E., & Hoste, V. (2018). Semeval-2018 task 3: Irony detection in english tweets. Proceedings of the 12th international workshop on semantic evaluation, 39–50.

Helal, N. A., Hassan, A., Badr, N. L., & Afify, Y. M. (2024). A contextual-based approach for sarcasm detection. Scientific Reports, 14(1), 15415.

Husein, A. M., Livando, N., Andika, A., Chandra, W., & Phan, G. (2023). Sentiment Analysis of Hotel Reviews on Tripadvisor with LSTM And ELECTRA. Sinkron: jurnal dan penelitian teknik informatika, 7(2), 733-740.

Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Applied Soft Computing, 91, 106198. https://doi.org/10.1016/J.ASOC.2020.106198

Jain, D. K., Kumar, A., & Sangwan, S. R. (2022). TANA: The amalgam neural architecture for sarcasm detection in indian indigenous language combining LSTM and SVM with word-emoji embeddings. Pattern Recognition Letters, 160, 11–18. https://doi.org/10.1016/J.PATREC.2022.05.026

Jain, T., Agrawal, N., Goyal, G., & Aggrawal, N. (2018). Sarcasm detection of tweets: A comparative study. 2017 10th International Conference on Contemporary Computing, IC3 2017, 2018-January. https://doi.org/10.1109/IC3.2017.8284317

Jamil, R., Ashraf, I., Rustam, F., Saad, E., Mehmood, A., & Choi, G. S. (2021). Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model. PeerJ Computer Science, 7. https://doi.org/ 10.7717/peerj-cs.645

Jansi, K. R. (2018). An extensive Survey on Sarcasm Detection Using Various Classifiers.International Journal of Pure and Applied Mathematics, 119(12).

Javdan, S., & Minaei-Bidgoli, B. (2020). Applying Transformers and Aspect-based Sentiment Analysis approaches on Sarcasm Detection, 67–71. https://doi.org/10.18653/ V1/2020.FIGLANG-1.9

Jiang, Z.-H., Yu, W., Zhou, D., Chen, Y., Feng, J., & Yan, S. (2020). Convbert: Improving bert with span-based dynamic convolution. Advances in Neural Information Processing Systems, 33, 12837–12848.

Joshi, A., Sharma, V., & Bhattacharyya, P. (2015). Harnessing context incongruity for sarcasm detection. ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association forComputational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, 2. https://doi.org/10.3115/v1/p15-2124

Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P., & Carman, M. (2016). Are word embedding based features useful for sarcasm detection? EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings. https://doi.org/10.18653/v1/d16-1104

Junaid, T., Sumathi, D., Sasikumar, A. N., Suthir, S., Manikandan, J., Khilar, R., Kuppusamy, P. G., & Raju, M. J. (2022). A comparative analysis of transformer based models for figurative language classification. Computers and Electrical Engineering, 101. https://doi.org/10.1016/j.compeleceng.2022.108051

Justo, R., Corcoran, T., Lukin, S. M., Walker, M., & Torres, M. I. (2014). Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. Knowledge-Based Systems, 69. https://doi.org/10.1016/j.knosys.2014.05.021

Kader, F. B., Nujat, N. H., Sogir, T. B., Kabir, M., Mahmud, H., & Hasan, K. (2022). Computational Sarcasm Analysis on Social Media: A Systematic Review. arxiv.org. https://arxiv.org/abs/2209.06170

Kalaivani, A., & Thenmozhi, D. (2020). Sarcasm identification and detection in conversion context using BERT. Proceedings of the Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.18653/v1/P17

Kandasamy, I., Vasantha, W. B., Obbineni, J. M., & Smarandache, F. (2020). Sentiment analysis of tweets using refined neutrosophic sets. Computers in Industry, 115. https://doi.org/10.1016/j.compind.2019.103180

Kara, B. K., & Karagoz, P. (2025, June). Data Augmentation with Translation for Turkish Irony Detection. In 2025 33rd Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Karabaş, A., & Dırı, B. (2020, October). Irony Detection with Deep Learning in Turkish Microblogs. In 2020 28th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Kaya, S., & Alatas, B. (2022). Sarcasm Detection with A New CNN+BiLSTM Hybrid Neural Network and BERT Classification Model. International Journal of Advanced Networking and Applications, 14. https://doi.org/10.35444/ijana.2022.14304

Khodak, M., Saunshi, N., & Vodrahalli, K. (2017). A large self-annotated corpus for sarcasm. arXiv preprint arXiv:1704.05579.

Koroteev, M. V. (2021). BERT: A review of applications in natural language processing and understanding. arXiv 2021. arXiv preprint arXiv:2103.11943.

Kumar, A., Narapareddy, V. T., Gupta, P., Srikanth, V. A., Neti, L. B. M., & Malapati,

A. (2020a). Adversarial and Auxiliary Features-Aware BERT for Sarcasm Detection. ACM International Conference Proceeding Series. https://doi.org/10.1145/ 3430984.3431024

Kumar, A., Narapareddy, V. T., Srikanth, V. A., Malapati, A., & Neti, L. B. M. (2020b). Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM. IEEE Access, 8. https://doi.org/10.1109/ACCESS.2019.2963630

Kumar, H. M., & Harish, B. S. (2018). Sarcasm classification: A novel approach by using Content Based Feature Selection Method. Procedia Computer Science, 143. https://doi.org/10.1016/j.procs.2018.10.409

Lad, R. (2023). Sarcasm Detection in English and Arabic Tweets Using Transformer Models. https://digitalcommons.dartmouth.edu/cs_senior_theses/20/

Lai, M., Bosco, C., Patti, V., & Virone, D. (2015). Debate on political reforms in Twitter: A hashtag-driven analysis of political polarization. Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015. https://doi.org/10.1109/DSAA.2015.7344884

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS. 8th International Conference on Learning Representations, ICLR 2020.

Lee, C. J., & Katz, A. N. (1998). The Differential Role of Ridicule in Sarcasm and Irony. Metaphor and Symbol, 13. https://doi.org/10.1207/s15327868ms1301_1

Li, G., Lin, F., Chen, W., & Liu, B. (2022). Affection Enhanced Relational Graph Attention Network for Sarcasm Detection. Applied Sciences (Switzerland), 12(7). https://doi. org/10.3390/app12073639

Li, J., Pan, H., Lin, Z., Fu, P., & Wang, W. (2021). Sarcasm Detection with Commonsense Knowledge. IEEE/ACM Transactions on Audio Speech and Language Processing,29. https://doi.org/10.1109/TASLP.2021.3120601

Li, J., Zhu, Y., & Sun, K. (2024). A novel iteration scheme with conjugate gradient for faster pruning on transformer models. Complex and Intelligent Systems, 10, 7863– 7875. https://doi.org/10.1007/S40747-024-01595-W/TABLES/3

Li, Q., Li, Z., Liu, W., He, X., & Pan, Y. (2025). Sarcasm-GPT: advancing sarcasm detection with large language models. The Computer Journal, bxaf055.

Liu, H., Wei, R., Tu, G., Lin, J., Liu, C., & Jiang, D. (2024). Sarcasm driven by sentiment: A sentiment-aware hierarchical fusion network for multimodal sarcasm detection. Information Fusion, 108, 102353.

Liu, H., Wang, W., & Li, H. (2022). Towards multi-modal sarcasm detection via hierarchical congruity modeling with knowledge enhancement. arXiv preprint arXiv:2210.03501.

Liu, L., Chen, X., & He, B. (2020). End-to-End Multi-task Learning for Allusion Detection in Ancient Chinese Poems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12275 LNAI. https://doi.org/10.1007/978-3-030-55393-7_27

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., & Allen, P. G. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://arxiv.org/abs/1907.11692v1

Liu, Y., Wang, Y., Sun, A., Meng, X., Li, J., & Guo, J. (2022). A Dual-Channel Framework for Sarcasm Recognition by Detecting Sentiment Conflict. Findings

of the Association for Computational Linguistics: NAACL 2022 - Findings. https://doi.org/ 10.18653/v1/2022.findings-naacl.126

Lukin, S., & Walker, M. (2017). Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue. arXiv preprint arXiv:1708.08572.

Lunando, E., & Purwarianti, A. (2013). Indonesian social media sentiment analysis with sarcasm detection. 2013 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2013. https://doi.org/10.1109/ICACSIS.2013.6761575

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

Maity, K., Jha, P., Saha, S., & Bhattacharyya, P. (2022). A multitask framework for sentiment, emotion and sarcasm aware cyberbullying detection from multi-modal code-mixed memes. Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, 1739–1749.

Maladry, A., Lefever, E., Hee, C. V., & Hoste, V. (2023). The limitations of irony detection in Dutch social media. Language Resources and Evaluation. https://doi.org/10. 1007/s10579-023-09656-1

Malik, M., Tomás, D., & Rosso, P. (2023). How challenging is multimodal irony detection?. In International Conference on Applications of Natural Language to Information Systems (pp. 18-32). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35320-8

Meaney, J.-A., Wilson, S., Chiruzzo, L., Lopez, A., & Magdy, W. (2021). SemEval 2021 task 7: HaHackathon, detecting and rating humor and offense. Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), 105–119.

Mehndiratta, P., & Soni, D. (2019a). Identification of Sarcasm in Textual Data: A Comparative Study. Journal of Data and Information Science, 4. https://doi.org/10.2478/ jdis-2019-0021

Mehndiratta, P., & Soni, D. (2019b). Identification of sarcasm using word embeddings and hyperparameters tuning. Journal of Discrete Mathematical Sciences and Cryptography, 22. https://doi.org/10.1080/09720529.2019.1637152

Meriem, A. B., Hlaoua, L., & Romdhane, L. B. (2021). A fuzzy approach for sarcasm detection in social networks. Procedia Computer Science, 192, 602–611. https://doi.org/10.1016/J.PROCS.2021.08.062

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. Ieee Access, 10, 99129–99149.

Mienye, I. D., & Swart, T. G. (2025). Ensemble Large Language Models: A Survey. Information, 16, 688.

Mishra, A., Kanojia, D., & Bhattacharyya, P. (2016). Predicting readers' sarcasm understandability by modeling gaze behavior. Proceedings of the AAAI Conference on Artificial Intelligence, 30.

Misra, R. (2022). News headlines dataset for sarcasm detection. arXiv preprint arXiv:2212.06035.

Misra, R., & Arora, P. (2023). Sarcasm detection using news headlines dataset. AI Open, 4, 13–18. https://doi.org/10.1016/J.AIOPEN.2023.01.001

Moffitt, B. (2016). The global rise of populism: Performance, political style, and representation. Stanford University Press.

Mohammed, P., Eid, Y., Badawy, M., & Hassan, A. (2019). Evaluation of different sarcasm detection models for arabic news headlines. In International Conference on Advanced Intelligent Systems and Informatics (pp. 418-426). Cham: Springer International Publishing.

Mukherjee, S., & Bala, P. K. (2017). Detecting sarcasm in customer tweets: An NLP based approach. Industrial Management and Data Systems, 117(6). https://doi.org/10. 1108/IMDS-06-2016-0207

Nagwanshi, P., & Madhavan, C. E. V. (2014). Sarcasm detection using sentiment and semantic features. KDIR 2014 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval. https://doi.org/10.5220/0005153504180424

Nayak, D. K., & Bolla, B. K. (2022). Efficient deep learning methods for sarcasm detection of news headlines. In Machine Learning and Autonomous Systems: Proceedings of ICMLAS 2021 (pp. 371-382). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-7996-4

Naz, F., Kamran, M., Mehmood, W., Khan, W., Alkatheiri, M. S., Alghamdi, A. S., & Alshdadi, A. A. (2019). Automatic identification of sarcasm in tweets and customer reviews. Journal of Intelligent and Fuzzy Systems, 37. https://doi.org/10.3233/ JIFS-190596

Okimoto, Y., Suwa, K., Zhang, J., & Li, L. (2021). Sarcasm Detection for Japanese Text Using BERT and Emoji. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12923 LNCS.

Onan, A., & Toçoğlu, M. A. (2020). Satire identification in Turkish news articles based on ensemble of classifiers. Turkish Journal of Electrical Engineering and Computer Sciences, 28(2), 1086-1106.

Oraby, S., Harrison, V., Reed, L., Hernandez, E., Riloff, E., & Walker, M. (2017). Creating and characterizing a diverse corpus of sarcasm in dialogue. arXiv preprint arXiv:1709.05404.

Ortega-Bueno, R., Rosso, P., & Fersini, E. (2023). Cross-Domain and Cross-Language Irony Detection: The Impact of Bias on Models' Generalization. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13913 LNCS.

Özturk, A. U., Cemek, Y., & Karagoz, P. (2021). IronyTR: Irony detection in Turkish informal texts. International Journal of Intelligent Information Technologies (IJIIT), 17(4), 1-18.

Özturk, A. U. (2024). Designing and debiasing binary classifiers for irony and satire detection [Yüksek lisans tezi, Orta Doğu Teknik Üniversitesi]. YÖK Tez Merkezi. https://tez.yok.gov.tr

Özkan, M., & Kar, G. (2022). Türkçe Dilinde Yazılan Bilimsel Metinlerin Derin Öğrenme Tekniği Uygulanarak Çoklu Sınıflandırılması. Mühendislik Bilimleri ve Tasarım Dergisi, 10. https://doi.org/10.21923/jesd.973181

Özyetgin, A. M. (2006). Tarihten Bugüne Türk Dili Alani.

Pandey, R., & Singh, J. P. (2023). BERT-LSTM model for sarcasm detection in code-mixed social media post. Journal of Intelligent Information Systems, 60. https://doi.org/10.1007/s10844-022-00755-z

Parameswaran, P., Trotman, A., Liesaputra, V., & Eyers, D. (2021). BERT's The Word: Sarcasm Target Detection using BERT. ALTA 2021 - Proceedings of the 19th Workshop of the Australasian Language Technology Association.

Parde, N., & Nielsen, R. (2018). Detecting Sarcasm is Extremely Easy ;-). https://doi.org/ 10.18653/v1/w18-1303

Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2016). A deeper look into sarcastic tweets using deep convolutional neural networks. COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers.

Potamias, R. A., Siolas, G., & Stafylopatis, A. (2019). A robust deep ensemble classifier for figurative language detection. In International Conference on Engineering Applications of Neural Networks (pp. 164-175). Cham: Springer International Publishing.

Potamias, R. A., Siolas, G., & Stafylopatis, A. G. (2020). A transformer-based approach to irony and sarcasm detection. Neural Computing and Applications, 32. https://doi.org/10.1007/s00521-020-05102-3

Ptáček, T., Habernal, I., & Hong, J. (2014). Sarcasm detection on Czech and English twitter. COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers.

Qi, F., Yang, C., Liu, Z., Dong, Q., Sun, M., & Dong, Z. (2019). Openhownet: An open sememe-based lexical knowledge base. arXiv preprint arXiv:1901.09957.

Qiao, Y., Jing, L., Song, X., Chen, X., Zhu, L., & Nie, L. (2023). Mutual-Enhanced Incongruity Learning Network for Multi-Modal Sarcasm Detection. Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023, 37. https://doi.org/10.1609/aaai.v37i8.26138

Qin, L., Huang, S., Chen, Q., Cai, C., Zhang, Y., Liang, B., Che, W., & Xu, R. (2023). MMSD2. 0: Towards a reliable multi-modal sarcasm detection system. arXiv preprint arXiv:2307.07135.

Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016). Election result prediction using Twitter sentiment analysis. In 2016 international conference on inventive computation technologies (ICICT) (Vol. 1, pp. 1-5). IEEE.

Ravi, K., & Ravi, V. (2017). A novel automatic satire and irony detection using ensembled feature selection and data mining. Knowledge-Based Systems, 120.

Ravi, K., & Ravi, V. (2018). Irony detection using neural network language model, psycholinguistic features and text mining. In 2018 IEEE 17th international conference on cognitive informatics & cognitive computing (ICCI* CC) (pp. 254-260). IEEE.

Ren, L., Xu, B., Lin, H., Liu, X., & Yang, L. (2020). Sarcasm Detection with Sentiment Semantics Enhanced Multi-level Memory Network. Neurocomputing, 401, 320–326.

Reyes, A., & Rosso, P. (2012). Making objective decisions from subjective data: Detecting irony in customer reviews. Decision Support Systems, 53, 754–760.

Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. Data & Knowledge Engineering, 74, 1–12.

Richardson, L. (2025). Beautiful Soup dokümantasyonu [Yazılım]. Erişim tarihi: 15 Ocak 2025, https://www.crummy.com/software/BeautifulSoup/

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. Proceedings of the 2013 conference on empirical methods in natural language processing, 704–714.

Rohanian, O., Taslimipoor, S., Evans, R., & Mitkov, R. (2018). WLV at SemEval-2018 Task 3: Dissecting Tweets in Search of Irony. NAACL HLT 2018 - International Workshop on Semantic Evaluation, SemEval 2018 - Proceedings of the 12th Workshop, 553–559. https://doi.org/10.18653/V1/S18-1090

Rosenthal, S., Farra, N., & Nakov, P. (2019). SemEval-2017 task 4: Sentiment analysis in Twitter. arXiv preprint arXiv:1912.00741.

Salim, S. S., Ghanshyam, A. N., Ashok, D. M., Mazahir, D. B., & Thakare, B. S. (2020). Deep LSTM-RNN with word embedding for sarcasm detection on Twitter. In 2020 international conference for emerging technology (INCET) (pp. 1-4). IEEE.

Samonte, M. J. C., Dollete, C. J. T., Capanas, P. M. M., Flores, M. L. C., & Soriano, C. B. (2018). Sentence-level sarcasm detection in English and Filipino tweets. ACM International Conference Proceeding Series.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. https://arxiv.org/abs/1910.01108v4

Saroj, A., & Pal, S. (2024). Ensemble-based domain adaptation on social media posts for irony detection. Multimedia Tools and Applications, 83(8), 23249-23268.

Savini, E., & Caragea, C. (2022). Intermediate-Task Transfer Learning with BERT for Sarcasm Detection. Mathematics, 10. https://doi.org/10.3390/math10050844

Schifanella, R., De Juan, P., Tetreault, J., & Cao, L. (2016). Detecting sarcasm in multimodal social platforms. MM 2016 - Proceedings of the 2016 ACM Multimedia Conference. https://doi.org/10.1145/2964284.2964321

Schweter, S. (2020). BERTurk - BERT models for Turkish. https://doi.org/10.5281/zenodo.3770924

Sekharan, S. C., Vadivu, G., & Rao, M. V. (2018). A Comprehensive Study On Sarcasm Detection Techniques In Sentiment Analysis. International Journal of Pure and Applied Mathematics, 118.

Sharma, C., Paka, W., Scott, D. B., Das, A., Poria, S., Chakraborty, T., & Gambäck, B. (2020). Task report: Memotion analysis 1.0@ semeval 2020: The visuo-lingual metaphor. Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020), Barcelona, Spain, Sep. Association for Computational Linguistics, 759–773.

Sharma, D. K., Singh, B., Agarwal, S., Kim, H., & Sharma, R. (2022). Sarcasm detection over social media platforms using hybrid auto-encoder-based model. Electronics, 11(18), 2844.

Sharma, Y., & Mandalam, A. V. (2021). Irony Detection in Non-English Tweets. 2021 6th International Conference for Convergence in Technology, I2CT 2021.

Shmueli, B., Ku, L.-W., & Ray, S. (2020). Reactive supervision: A new method for collecting sarcasm data. arXiv preprint arXiv:2009.13080.

Shrikhande, P., Setty, V., & Sahani, A. (2020). Sarcasm Detection in Newspaper Headlines. 2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings.

Son, L. H., Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term

memory model with convolution network. IEEE Access, 7, 23319–23328. https://doi.org/10.1109/ACCESS.2019.2899260

Srivastava, H., Varshney, V., Kumari, S., & Srivastava, S. (2020). A Novel Hierarchical BERT Architecture for Sarcasm Detection. https://doi.org/10.18653/v1/2020.figlang-1.14

Stranisci, M., Bosco, C., Farías, D. I. H., & Patti, V. (2016). Annotating sentiment and irony in the online Italian political debate on LaBuonaScuola. Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016.

Subramanian, J., Sridharan, V., Shu, K., & Liu, H. (2019). Exploiting emojis for sarcasm detection. In International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation (pp. 70-80). Cham: Springer International Publishing.

Sukhavasi, V., & Dondeti, V. (2025). Multi-modal sarcasm detection using ensemble net model. Knowledge and Information Systems, 67, 403–425.

Sulis, E., Irazú Hernández Farías, D., Rosso, P., Patti, V., & Ruffo, G. (2016). Figurative messages and affect in Twitter: Differences between irony, sarcasm and not [New Avenues in Knowledge Bases for Natural Language Processing]. Knowledge-Based Systems, 108, 132–143.

Sundararajan, K., & Palanisamy, A. (2020). Multi-rule-based ensemble feature selection model for sarcasm type detection in Twitter. Computational Intelligence and Neuroscience, 2020. https://doi.org/10.1155/2020/2860479

Swami, S., Khandelwal, A., Singh, V., Akhtar, S. S., & Shrivastava, M. (2018). A corpus of English-Hindi code-mixed tweets for sarcasm detection. 2018. Centre for Language Technologies Research Centre.

Talafha, B., Za'Ter, M. E., Suleiman, S., Al-Ayyoub, M., & Al-Kabi, M. N. (2021). Sarcasm detection and quantification in Arabic tweets. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 1121-1125). IEEE.

Tasar, B., Ozcan, İ., Tatar, A. B., & Yakut, O. (2019). Destek Vektör Makinasi Algoritması ile Kalp Hastalıklarının Tahmini. Computer Science, 4, 74–79. http://dergipark. org.tr/tr/download/article-file/833627

Tasnia, R., Ayman, N., Sultana, A., Chy, A. N., & Aono, M. (2023). Exploiting stacked embeddings with LSTM for multilingual humor and irony detection. Social Network Analysis and Mining, 13. https://doi.org/10.1007/s13278-023-01049-0

Taşlıoğlu, H. (2014). Ironi detection on turkish microblog texts [Yüksek lisans tezi, Orta Doğu Teknik Üniversitesi]. YÖK Tez Merkezi. https://tez.yok.gov.tr

Taşlıoğlu, H., & Karagoz, P. (2017, April). Irony detection on microposts with limited set of features. In Proceedings of the Symposium on Applied Computing (pp. 1076-1081).

Tay, Y., Tuan, L. A., Hui, S. C., & Su, J. (2018). Reasoning with sarcasm by reading in-between. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 1. https://doi.org/10.18653/v1/p18-1093

Teodorescu, V., & Obreja Brașoveanu, L. (2025). Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost. Computation, 13(5), 127.

Thakur, S., Singh, S., & Singh, M. (2018). Detecting sarcasm in text. In International Conference on Intelligent Systems Design and Applications (pp. 996-1005). Cham: Springer International Publishing.

Toçoğlu, M. A., & Onan, A. (2019, August). Satire detection in Turkish news articles: a machine learning approach. In International Conference on Big Data Innovations and Applications (pp. 107-117). Cham: Springer International Publishing.

Tomás, D., Ortega-Bueno, R., Zhang, G., Rosso, P., & Schifanella, R. (2023). Transformer-based models for multimodal irony detection. Journal of Ambient Intelligence and Humanized Computing, 14(6), 7399-7410.

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.

Vinoth, D., & Prabhavathy, P. (2022). An intelligent machine learning-based sarcasm detection and classification model on social networks. Journal of Supercomputing,78. https://doi.org/10.1007/s11227-022-04312-x

Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., & King, J. (2012). A Corpus for Research on Deliberation and Debate. LREC, 12, 812–817.

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. Decision Support Systems, 57. https://doi.org/10.1016/ j.dss.2013.08.002

Wang, L., Cao, Q., Zhang, Z., Mirjalili, S., & Zhao, W. (2022). Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems. Engineering Applications of Artificial Intelligence, 114, 105082. https://doi.org/10.1016/J.ENGAPPAI.2022.105082

Wen, Z., Gui, L., Wang, Q., Guo, M., Yu, X., Du, J., & Xu, R. (2022). Sememe knowledge and auxiliary information enhanced approach for sarcasm detection. Information Processing & Management, 59, 102883. https://doi.org/10.1016/J.IPM.2022.102883

Xu, Z., Gong, L., Ke, G., He, D., Zheng, S., Wang, L., Bian, J., & Liu, T.-Y. (2020). MC-BERT: Efficient Language Pre-Training via a Meta Controller. https://arxiv.org/ abs/2006.05744v2

Yabanova, İ., & Yumurtacı, M. (2018). Destek vektör makineleri kullanarak dinamik yumurta ağırlıklarının sınıflandırılması. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 33, 393–402.

Yakut, E., Elmas, B., & Yavuz, S. (2014). Yapay sinir ağlari ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 19, 139–157. http://dergipark.org.tr/tr/download/ article-file/194261

Yan, Y., Zhang, B. W., Ding, G., Li, W., Zhang, J., Li, J. J., & Gao, W. (2023). O 2 - Bert: Two-Stage Target-Based Sentiment Analysis. Cognitive Computation. https://doi.org/10.1007/s12559-023-10191-y

Yang, Y., Lv, H., & Chen, N. (2023). A survey on ensemble learning under the era of deep learning. Artificial Intelligence Review, 56, 5545–5589.

Yao, F., Sun, X., Yu, H., Zhang, W., Liang, W., & Fu, K. (2023). Mimicking the Brain's Cognition of Sarcasm from Multidisciplines for Twitter Sarcasm Detection. IEEE Transactions on Neural Networks and Learning Systems, 34. https://doi.org/10. 1109/TNNLS.2021.3093416

Yue, T., Mao, R., Wang, H., Hu, Z., & Cambria, E. (2023). KnowleNet: Knowledge fusion network for multimodal sarcasm detection. Information Fusion, 100. https://doi. org/10.1016/j.inffus.2023.101921

Yunitasari, Y., Musdholifah, A., & Sari, A. K. (2019). Sarcasm Detection For Sentiment Analysis in Indonesian Tweets. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 13. https://doi.org/10.22146/ijccs.41136

Zanchak, M., Vysotska, V., & Albota, S. (2021). The sarcasm detection in news headlines based on machine learning technology. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2. https://doi. org/10.1109/CSIT52700.2021.9648659

Zeng, B., Yang, H., Xu, R., Zhou, W., & Han, X. (2019). LCF: A Local context focus mechanism for aspect-based sentiment classification. Applied Sciences (Switzerland),9(16). https://doi.org/10.3390/app9163389

Zhang, L., Zhao, X., Song, X., Fang, Y., Li, D., & Wang, H. (2022). A Novel Chinese Sarcasm Detection Model Based on Retrospective Reader. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13142 LNCS. https://doi.org/10.1007/978-3-030-98355-023

Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wu, F., & Wang, G. (2023). Instruction Tuning for Large Language Models: A Survey. https://arxiv.org/abs/2308.10792

Zhang, Y., Zhu, G., Ding, Y., Wei, Z., Chen, L., & Li, K.-C. (2025). A progressive interaction model for multimodal sarcasm detection. The Journal of Supercomputing, 81, 1–28.

Zhong, W., Zhang, Z., Wu, Q., Xue, Y., & Cai, Q. (2024). A semantic enhancement framework for multimodal sarcasm detection. Mathematics, 12(2), 317.

Zhu, N., & Wang, Z. (2020). The paradox of sarcasm: Theory of mind and sarcasm use in adults. Personality and Individual Differences, 163. https://doi.org/10.1016/j. paid.2020.110035

Zuhri, A. T., & Sagala, R. W. (2022). Irony and sarcasm detection on public figure speech. J. Elem. School Educ., 1(1), 41-45.